DOES ACTIVE LIVING AFFECT SCHOOL PERFORMANCE?

Bart H.H. Golsteyn[a], Maria W.J. Jansen[b,c], Dave H.H. Van Kann[c,d,e], Annelore M.C. Verhagen[a,*]

[a] Department of Economics, Maastricht University

[b] Department of Health Services Research, Care and Public Health Research Institute, Maastricht University

[c] Academic Collaborative Centre for Public Health Limburg, Public Health Services

[d] Department of Health Promotion, Care and Public Health Research Institute, Maastricht University

[e] School of Sport Studies, Fontys University of Applied Sciences

APRIL, 2018

Abstract

This paper investigates whether stimulating physical activity in everyday life affects primary school performance. We use data from the Active Living field experiment, which aims to increase active transportation and active play among 8-12 year-olds living in low-SES areas in the Netherlands. Difference-in-differences estimations reveal that while Active Living increases time spent on physical activity during school hours, it negatively affects school performance, especially among the worst-performing students. Our results suggest that, although physical exercise may be beneficial for health, stimulating active play may have negative effects on educational outcomes.

Keywords: Health behavior; Field experiment; Education; Physical activity.

JEL codes: I12; C93; I20.

---

[*] Corresponding author (e-mail: a.verhagen@maastrichtuniversity.nl, phone: +31 (0)43 388 3620). Author names are in alphabetical order. We are grateful to all schools and children participating in this study, and all Maastricht University colleagues involved in the data collection. A previous version of the paper was presented at the Applied Economics Lunch Seminar at Paris School of Economics, the Learning and Work seminar at Maastricht University, the Netherlands Economists Day 2017, the 29th annual conference of the European Association of Labour Economists, and the 31st Annual Conference of the European Society for Population Economics. We thank participants of these conferences and seminars, as well as Eric Bonsang, Lex Borghans, Elizabeth Cascio, Stefano DellaVigna, Thomas Dohmen, Ilyana Kuziemko, Michael Lechner, and Bas Ter Weel, for valuable comments and suggestions. This study was funded by the Netherlands Organization for Scientific Research (VIDI grant 452-16-006), and the Netherlands Organization for Health Research and Development (ZonMW), Project Number 200130003.

I. Introduction

Due to the global increase of childhood obesity and sedentary behavior, organizations and governments advocate increasing the amount of time elementary school children spend on physical activity.[2] While the benefits of physical activity for physical and mental health are well documented, the current literature is still ambiguous about the causal effects on educational outcomes. From a physiological viewpoint, physical activity is expected to positively affect cognition (and thereby school performance),[3] but from an economic perspective, the effect of increasing physical activity on school performance is theoretically ambiguous. Even if physical activity has positive effects on children's cognitive abilities, increasing time spent on doing sports may crowd-out time investments in other potentially beneficial activities, such as studying, active play, or focus and attention during instruction time.[4] Physical activity may, however, also crowd-out potentially harmful activities such as smoking or watching TV, leading to a zero or positive effect on school performance.[5] The net effect on school performance thus depends on the relative gains (or harm) from the activity that was crowded-out, compared to the gains from physical activity.

This paper investigates with a field experiment whether stimulating informal physical activity such as active play, affects primary school performance. The experiment is organized in low-

---

[2] See, e.g., World Health Organization (2010); Centers for Disease Control and Prevention (2011); European Commission (2014).

[3] The medical literature shows that physical activity can positively affect cognitive functioning through increased blood and oxygen flow to the brain, decreased stress and improved mood due to increased levels of norepinephrine and endorphins, and through increased growth factors that help to create new nerve cells and support synaptic plasticity (Singh et al. 2012). Erickson, Hillman and Kramer (2015), for example, find that more active children show greater hippocampal and basal ganglia volume, greater white matter integrity, elevated and more efficient patterns of brain activity, and superior cognitive performance and scholastic achievement.

[4] As Lizandra et al. (2016) point out, sedentary activities could either be academic (e.g. doing homework, reading or studying), technological-based (e.g. playing video games or watching TV) or social-based (e.g. sitting with friends or chatting via social networks), and each of these behaviors has a different expected effect on academic performance.

[5] Another potential mechanism that has not yet received much attention in the literature is that physical activity could increase physical or cognitive fatigue. If children and their parents/teachers devote more time and energy to sports or play and less to school work, the effect of physical activity on school performance could be negative.

SES areas of the Netherlands. We analyze whether the experiment affects physical activity and school performance.

Analyzing the effect of physical activity on school performance entails the risk of reverse causality or selection based on unobserved factors. For example, an unobserved deterioration of health or decrease in study motivation may cause a simultaneous decrease in school performance and time spent on physical activity. An example of reverse causality is that children with decreasing school performances may be encouraged to spend more time doing homework (a sedentary activity). Previous attempts to overcome these problems include adopting an individual-fixed-effects approach (Lipscomb 2007; Rees and Sabia 2010), and selection-on-observables or matching (Pfeifer and Cornelissen 2010; Cabane, Hille, and Lechner 2016; Felfe, Lechner, and Steinmayr 2016). Although these studies are interesting contributions to the literature, both approaches may still suffer from an omitted variables bias. Some studies therefore (additionally) adopt an instrumental variables approach, in which students' sports participation is instrumented by their height (Barron, Ewing, and Waddell 2000; Rees and Sabia 2010; Pfeifer and Cornelissen 2010), or by certain parental characteristics such as income or school characteristics such as school size and the books-per-student ratio (Barron, Ewin, and Waddell 2000). The problem with these identification strategies is that it is questionable whether the exclusion restriction holds. Stevenson (2010) and Dills, Morgan and Rotthoff (2011) are exceptions to use changes in state-level policies to instrument for changes in sports participation at school.[6]

Our strategy to overcome the before mentioned problems of reverse causality or selection based on unobserved factors is to organize the *Active Living* program: a field experiment among

[6] Stevenson (2010) finds a positive effect of increased high school athletic participation on education and labor market outcomes for women. Dills, Morgan and Rotthoff (2011) do not find any significant effects of increased recess or physical education time on school performance.

8-12 year olds, which aims to increase physical activity and decrease sedentary behavior through active transportation and active play at school and during leisure time. The interventions, which were implemented between April 2013 and June 2014, did not affect school time or the structure or content of classes. Pre and post-treatment data regarding school performance and time spent on physical activity were collected in 10 treatment schools and 11 control schools. Our measures of school performance are based on nationally standardized language and math/calculus tests, with a grading scheme that is the same across all schools. Physical activity data were collected by means of accelerometers that were worn for five consecutive days both at pre and post-treatment measurement. We use a difference-in-differences technique to estimate the effects of Active Living on school performance and time spent on physical activity. By including child fixed-effects, we control for any time-invariant individual characteristics and take into account that unbalanced panel data may otherwise bias our results. We allow for heterogeneous treatment effects by analyzing the effects of Active Living for a moving window of 200 individuals across the school performance distribution in the pre-treatment year 2012/2013.

The results indicate that Active Living causes a significant decrease in overall school performance. Our most conservative estimate of the average treatment effect is 5.9 percent of a standard deviation. The results are similar when we estimate the effect on language and math tests separately. Analyses across the pre-treatment school performance distribution indicate that the negative effect on school performance is strongest among the worst-performing students. Several robustness checks provide strong indications that the effects on school performance are indeed due to the Active Living interventions in the treatment schools. Moreover, since Active Living causes a significant increase in time spent on physical activity during school time of 0.34 standard deviations (9.3 minutes per day), with no significant effect on leisure time activity, it is likely that physical activity at school is an important mechanism for the negative effect on school performance.

This paper contributes to the literature studying causal effects of physical activity interventions on educational outcomes. Earlier papers on the consequences of physical exercise have mostly shown positive effects on health (see, e.g., Strong et al. 2005; Cawley, Meyerhoefer, and Newhouse 2007; Cawley, Frisvold, and Meyerhoefer 2013) and on cognitive functioning, particularly in older adults (see, e.g., Colcolmbe and Kramer 2003; Hillman, Erickson, and Kramer 2008). As indicated earlier, a very limited set of papers analyzes the causal effect of physical activity on educational outcomes of younger individuals, but besides the remaining difficulties regarding the identification strategies, these studies focus on the effect of physical education classes, doing sports in clubs, or other forms of *formal* or professional sport activities (see, e.g., Barron, Ewing, and Waddell 2000; Eide and Ronan 2001; Lechner 2009; Pfeifer and Cornelissen 2010; Stevenson 2010; Cabane, Hille, and Lechner 2016). Our study contributes to this literature by being the first to analyze the effect of a field experiment which mainly focuses on encouraging *informal* physical activities such as active transportation and active play.

A first reason why the focus on informal physical activity is an important contribution to the literature is that being physically active entails more than doing physical exercise alone. During the course of a day, the amount of time spent on informal physical activity is (or can be) substantially larger than the amount of time spent on physical exercise. Secondly, there are several reasons why it is unclear whether we should expect the effects of stimulating informal physical activity to be similar (i.e., positive) to the effects of formal physical activity which are usually found. Firstly, since informal physical activities can be expected to be of lower intensity than doing sports, our study provides more insight into the dose-response relationship between physical activity and school performance. Previous research has shown that the positive relation between physical activity or health and educational outcomes may be mostly driven by moderate and vigorous-intensity physical activity (see, e.g. Singh et al. 2012; Van den Berg et al. 2016; Felez-Nobrega et al. 2017). Secondly, formal and informal physical activity may foster the development of different skills. Sports, and athletics in particular, teaches players competitiveness and perseverance; skills which could be beneficial for educational outcomes, but which are likely less pronounced (if at all present) in active play. Additionally, athletes may receive increased attention and encouragement from their teachers or parents, leading to increased self-esteem and increased peer pressure to succeed (Stevenson 2010). Developing these sports-related skills means that potential crowding-out effects will be less pronounced. In other words, if children usually reallocate study hours to physical activity, the effect of doing sports on school performance might have been negative, had it not been for the development of the additional (sports-related) skills. The absence of these additional skills when informal physical activity is stimulated, might then lead to a zero or negative effect on school performance.

Another important difference as compared to other physical activity intervention studies is that none of the interventions affected school time or the curriculum. Moreover, our physical activity data are collected by means of accelerometers that were worn throughout the day, which allows us to capture any substitution or complementarity effects between in and out of school physical activity, and between formal and informal forms of physical activity. This is an important benefit relative to other analyses, where the focus lies on (self-reported) measures of formal forms of physical activity.

This paper proceeds as follows. Section II gives the empirical strategy and discusses the field experiment. Section III describes the data, Section IV the results, and Section V concludes.

II. Empirical strategy

A. The Active Living program[7]

In order to stimulate physical activity (PA), we organized the *Active Living* program: a field experiment which took place between 2012 and 2014 in 21 primary schools located in low-SES areas in the Southern-Limburg region of the Netherlands (see Figure A1 in the Appendix for a map of the Active Living research area).[8] The target group of the experiment consists of children who were enrolled in grade 4 or 5 during the 2012/2013 school year, and transferred to grade 5 or 6 during the 2013/2014 school year (i.e., 8-12 year olds).

Ten primary schools agreed to participate in the experiment as a treatment school. Each treatment school was matched to a control school based on the level of neighborhood deprivation and the level of urbanization (urban versus rural). One treatment school was matched to an additional control school, because the school to which it was initially matched relocated during the 2012/2013 school year. After the pre-treatment PA measurement, no further changes were observed in this control school, making both control schools eligible for inclusion in this study. One treatment school was excluded from the data because it organized a sports day during the pre-treatment PA measurement, resulting in exceptionally high levels of PA that were not representative for children's average school days.

[7] The text in this section is partly taken from Van Kann et al. (2016). For a more detailed description of the study design of the Active Living program, its methods and implementation strategies, we refer to Van Kann et al. (2015).

[8] The Limburg region has about 609,000 inhabitants, and a population density of 922 inhabitants per square kilometer. This density is almost twice as high as the average population density in the Netherlands, which is 496 inhabitants per square kilometer. The region consists of 18 municipalities. The average disposable household income in the region is about 31,500 Euro per year, which is somewhat lower than the national average of 34,200 Euro per year (Statistics Netherlands 2013).

A working group was composed at each treatment school, consisting of representatives of the school, parents, municipal authorities and other stakeholders, and a public health services (PHS) employee as chair. The PHS employee was trained by the researchers in '(physical/social) environmental thinking' and evidence-based PA intervention opportunities. The working groups were responsible for choosing, designing and implementing intervention elements, but final plans were financially approved by the research team. Although sharing an overall scope, the intervention packages could differ in magnitude and design across treatment schools, depending on local needs.

All working groups received an intervention budget of 2000 euros at the start of the project. However, if intervention plans exceeded the budget, additional funding resources could be applied for at the research team and municipal authorities (for which a representative was a member of the working group). Additionally, the province of Limburg provided funding for organizing a sports day at all treatment schools, and several treatment schools organized charity events such as a charity run in order to raise money for new playground equipment. Treatment schools were allowed to start the implementation of interventions as soon as the pre-treatment PA measurement was conducted at their school (see Section III for more information on the PA data collection). All interventions were implemented before the second PA measurement in the spring of 2014. An overview of all interventions per treatment school is provided in Table 1.[9]

[Insert Table 1 Here]

Each intervention intends to decrease sedentary behavior or increase physical activity at school or in leisure time. It is important to note that none of the interventions affects school time, the structure or content of the curriculum, since otherwise this may affect school performance. For the same reason, interventions requiring structural changes in the school yard (e.g., installing

[9] See Table A1 in the Appendix for a description of each intervention.

new fixed equipment, a ball backstop, or sound equipment in the school yard) were implemented after school hours.

In order to decrease the probability that children from control schools are affected by the interventions (i.e., contamination of the control group), all interventions were implemented at the treatment schools or within an 800 meter radius around the schools (as illustrated in Figure A1 in the Appendix). In the Netherlands, the majority of the primary school children live within 800 meter distance from school (Statistics Netherlands 2014). Control schools did not implement any Active Living interventions.

B. Identifying the Active Living effect on school performance

After the data collection (see Section III for more information), we verified whether treatment and control schools are similar with respect to pre-treatment observables. Table A2 in the Appendix reveals that, before any interventions were implemented, treatment and control schools differed on various domains, the most important differences being that children in treatment schools performed significantly better on math tests and spent on average 15 minutes less per day on PA during school hours. This means that we cannot merely compare post-treatment outcomes between children in treatment and control schools, because post-treatment differences may be due to self-selection.

When pre-treatment differences between the treatment and control group exist, a difference-in-differences technique can be an alternative strategy for identifying treatment effects. The identifying assumption underlying the technique is that the initial differences between the treatment and control group would have remained similar if Active Living had not taken place.

The common trends in the pre-treatment years indicate that it is plausible that this assumption holds (see Figure 1, as well as Table A3 in the Appendix).[10]

[Insert Figure 1 Here]

In order to identify the effect of Active Living on school performance, we estimate the following model:

$$(1)\ Score_{ist} = \alpha_0 + \alpha_1 DiD_{st} + \alpha_2 Treat_s + \alpha_3 Post_t + \alpha_4 age_{ist} + \alpha_5 age^2{}_{ist} + \gamma_{1is} + \varepsilon_{1ist}$$

In this equation, *i* indexes individual pupil, *s* indexes school, and *t* indexes time. In the regressions, *Treat* is an indicator variable with value 1 if a pupil is in a school that belongs to the treatment group and 0 if (s)he is enrolled in one of the control schools.[11] *Post* is an indicator variable with value 1 for tests taken during the post-treatment year (2013/2014) and 0 for tests taken during the pre-treatment year (2012/2013).[12] We are interested in the difference-in-differences estimator α_1 for the *DiD* variable, i.e. the interaction of *Treat* and *Post*. Our baseline specification includes the *DiD*, *Treat* and *Post* variables. In our main specification, we also control for age and child fixed-effects γ.[13] By including age squared, we allow for nonlinear age effects. Individual-specific time-invariant characteristics, such as innate ability, gender or cohort (whether the child was enrolled in 4th or 5th grade during the pre-treatment year) are captured by including child fixed-effects. We thereby also take into account that unbalanced panel data may bias our results when those who are observed only in one time period have different characteristics from those who are observed in both time periods.

[10] Treating *Year* as a continuous variable instead of including year-fixed-effects yields the same results: differences between the treatment and control group do not change significantly in the pre-treatment years.

[11] There are no children in the sample who changed school between the pre and post-treatment year.

[12] Note that the interventions could be implemented *throughout* the 2013/2014 school year (until March-June 2014), which implies that some tests taken during this school year may be wrongly defined as post-treatment tests. However, since this would attenuate results, we consider this to be less of a problem.

[13] It is possible to include both age and year fixed-effects, because in our sample, age is measured in months when tests are taken, and test dates vary by year.

***Robust standard errors*.—**School performance may be correlated among pupils within the same school. Moreover, by construction, the difference-in-differences variable as well as the treatment and year indicator variables are constant within schools. Ignoring such within-group dependence can greatly underestimate the true standard errors. A common strategy to correct for this is to cluster the standard errors at the school level. However, asymptotic justification assumes that the number of clusters goes to infinity, whereas "only" 21 schools participated in the Active Living program. Previous research has demonstrated that in case of few clusters, robust standard errors can be substantially downward biased (see, e.g., Bell and MacCaffrey 2002; Bertrand, Duflo, and Mullainathan 2004; Donald and Lang 2007; Angrist and Pischke 2008; Cameron, Gelbach, and Miller 2008; Cameron and Miller 2015).

In order to overcome these problems, we use the method proposed by Donald and Lang (2007), which has been shown to work reasonably well in difference-in-differences settings. They suggest to base inference on a *t*-distribution with degrees of freedom equal to the number of clusters minus the number of variables that are constant within clusters, rather than the standard normal distribution. By doing so, one essentially recognizes that the fundamental unit of observation is a cluster and not an individual within a cluster. We verified that inference based on this approach is indeed most conservative compared to no clustering, clustering without adjusted critical *t*-values, or clustering with corrected standard errors based on the Moulton factor (as proposed by Angrist and Pischke 2008).[14] Furthermore, in order to avoid serial correlation (Bertrand, Duflo, and Mullainathan 2004), we compare the pre-treatment year (2012/2013) to the post-treatment year, instead of using test scores in all pre-treatment years as the baseline measure.

[14] In some cases, inference based on one of the other strategies yields slightly more conservative results, but in these cases this does not change the interpretation of the results, i.e. the rejection of the null-hypothesis.

C. Mechanism

We also investigate a mechanism through which Active Living might have an effect on school performance. The most likely mechanism is through increased time spent on PA. We identify the effect of Active Living on time spent on PA by estimating the following model:

$$(2)\ PA_{ist} = \beta_0 + \beta_1 DiD_{st} + \beta_2 Treat_s + \beta_3 Post_t + \boldsymbol{X'_{ist}}\beta + \gamma_{2is} + \varepsilon_{2ist}$$

The *DiD*, *Treat* and *Post* variables, as well as child fixed-effects γ, are identical to those in equation (1). ***X'*** is a vector of variables, including the time children wear the accelerometer per day (for more information on the accelerometers, see Section III) and several variables indicating the weather conditions during each PA measurement week (i.e., temperature, hours of sunshine, and rain fall). In order to allow for nonlinear relationships between these variables and time spent on PA, we also control for wearing time squared, temperature squared, and hours of sunshine squared. Because the study design for Active Living entails that the pre and post-treatment PA measurement are scheduled to take place around the same time for each school, age is approximately constant in the PA data once we control for year and child fixed-effects. Considering the Active Living interventions that are implemented, we expect to find that the difference-in-differences estimator β_1 is positive and significantly different from zero, especially if Active Living significantly affects school performance.

III. Data

The analyses are based on two data sets. The first contains data on individual test score histories from children in nearly all schools in the Southern Limburg region ("school performance data"). The second contains data on time spent on PA as measured by accelerometers ("PA data") worn by children in the 21 schools that participated in the Active Living program. We discuss these data sets in turn. See Figure 2 for a timeline of the data collection process.

[Insert Figure 2 Here]

A. School performance

The data on school performance are collected at over 200 schools in a cooperative project between elementary schools, school boards, municipalities and Maastricht University, called the *Onderwijs Monitor Limburg*. The 21 Active Living schools are included in these 200 schools. Data on the school performance of children in the schools that did not participate in the Active Living program as a treatment or control group are used in the robustness checks where we compare the school performance of children in the control group and treatment group to the school performance of children in non-Active Living schools.

The school performance data include a wide variety of standardized tests that children took between Kindergarten and grade 6 (i.e., age 4 to 12) in various domains such as reading, language, calculating and math. The tests are graded by the teachers, but with a grading scheme that is standardized across all schools in the Netherlands. Each subject has its own range of scores.[15] For each subject, there are two tests per grade with similar contents. The teacher can decide whether a child takes both tests and whether a child takes the same test more than once.

For the purpose of this study, we limit our sample to the target group of the Active Living program, i.e. children who were enrolled in grade 4 or 5 in 2012/2013, and in grade 5 or 6 in 2013/2014.[16] We exclude tests that were taken during the PA measurements to prevent reactivity effects. Because in each year, the last PA measurement day took place only one week before the

[15] For example, math test scores range from 0 to 168, whereas reading test scores range from -87 to 147.

[16] There are no children in our sample who repeated 4th grade in 2013/2014. Children who repeated 5th grade in 2013/2014, as well as those who skipped 5th grade and went directly from 4th grade in 2012/2013 to 6th grade in 2013/2014, are included in the analyses.

start of the summer holidays, we also exclude the few tests that were taken during this week. No tests are taken during the summer holidays. This means that we exclude all tests that were taken between the April 3rd and August 11th 2013, and between March 26th and August 24th 2014.[17]

Because children can take multiple tests of the same subject within a school year, we first aggregate children's subject-scores to the average within each school year. In order to create a comparable scale across subjects (which each have their own range of scores), we standardize children's average subject-scores across schools (Active Living schools as well as non-Active Living schools) and school years to a mean of zero and a standard deviation of one. We can then construct an overall school performance variable, by calculating children's average score across their standardized subject-scores within a school year. For interpretation purposes, we rescale this variable on a scale from 0 to 100. This is the dependent variable in our main analyses on school performance. We repeat this strategy for language and math tests separately.

B. Time spent on PA

The school performance data can be merged at the individual level with the PA data. Time spent on PA was measured during one week between April 3rd and June 24th 2013 and again during one week 12 months later (March 26th-July 7th 2014). Prior to the pre-treatment PA measurement, written parental consent to wear an accelerometer was obtained for 61.6% of the children in the target group. We verified that these children were not significantly different with respect to observable characteristics as compared to those for whom we did not receive parental consent to wear an accelerometer.[18] In the pre-treatment year, 791 children were fitted with an accelerometer which they were asked to wear on their waist for at least five consecutive days,

[17] The results from analyses including these tests remain qualitatively robust (results available upon request).
[18] Results are available upon request.

including a weekend day.[19] To correct for potential seasonal effects (see, e.g., Rich, Griffiths, and Dezateux 2012), every treatment-control school pair was assessed at the same dates.

PA data were collected using accelerometers (ActiGraph GT3X+; 30 Hz, 10-second epochs). The sensors in accelerometers measure acceleration in units of gravity on three axes: vertical, horizontal and perpendicular. The accelerometers run on a battery and are not turned on or off by the participants. They register acceleration continuously. In this study, we use Evenson's cut-offs (Evenson et al. 2008) to determine PA intensity levels: accelerations at more than 100 counts-per-minute are registered as time spent on PA and accelerations of 100 counts-per-minute or less are registered as time spent on sedentary behavior.[20] We make a distinction between sedentary behavior and not-wearing of the accelerometer based on Choi's classification criteria (Choi et al. 2011). We use the ActiLife software (version 6.10.4, ActiGraph, Pensacola, USA) to transform the raw data into data per half-hour.

School time PA is defined according to regular school times of Dutch primary schools, i.e. Mondays, Tuesdays, Thursdays and Fridays from 9AM until 3PM, and Wednesday from 9AM until noon. Leisure time is defined as the hours before and after school time and the weekends. PA data collected between 11PM and 6AM are excluded from the analyses, because these data are defined as sleeping hours.[21] We sum the PA data per daypart (school time/leisure time) for each child.

---

[19] Parental consent was asked to parents of 1,322 children, and obtained for 815 children (61.6%). Out of the 815 children for whom we obtained parental consent to wear an accelerometer, 791 were fitted with an accelerometer in the pre-treatment year (97%); 24 children did not wear their accelerometer.

[20] Counts-per-minute are a unit of activity that is commonly used in health sciences. Light intensity PA (e.g. walking, biking slowly, playing catch) is defined as 101–2,295 counts-per-minute, 2,296 – 4,011 counts-per-minute is defined as moderate intensity PA (e.g. brisk walking, jumping on a trampoline, recreational swimming), and ≥4,012 counts-per-minute is defined as vigorous intensity PA (e.g. running, jumping rope, swimming laps). Fewer than 100 counts-per-minute is defined as sedentary behavior (e.g. watching TV, gaming, or making homework) (Evenson et al. 2008).

[21] Usually, these data are already excluded because they are recognized as non-wearing time (Choi et al. 2011). Excluding the 11PM-6AM data reduces the number of child-hour observations by 1.91%.

We define a minimum amount of wearing time of the accelerometer, because we otherwise would have to assume that, if a child for example wore the accelerometer for only five minutes during an entire day, that those five minutes are representative for the child's activity levels throughout that day. We define minimal wear time as at least half the daily amount of school/leisure time. This means that, regarding school time PA, minimal wear time is defined as at least 180 minutes (3 hours) between 9AM and 3PM on Mondays, Tuesdays, Thursdays and Fridays, and at least 90 minutes (1.5 hours) between 9AM and noon on Wednesdays.[22] Minimal wear time for leisure time PA is defined as at least 330 minutes (5.5 hours) on Mondays, Tuesdays, Thursdays and Fridays, at least 420 minutes (7 hours) on Wednesdays, and at least 480 minutes (8 hours) on Saturdays and Sundays.[23]

We aggregate the children's daily time spent on PA (school time/leisure time) to the average amount of time spent on PA per measurement week. This means that in our analyses, PA is measured in average number of minutes per day (per measurement week).

C. Control variables

In the school performance data, age is registered as the average age (in months) on the day the tests were taken. In the PA data, age is registered as age (in months) during the PA measurement week. Both data sets include an indicator variable with value 1 if the child is a boy. We create an indicator variable with value 1 if the child was enrolled in grade 5 during the pre-treatment year, and 0 if in this school year, the child was enrolled in grade 4. As described earlier, wearing time of the accelerometer is registered in minutes. Our weather indicators are mean temperature, hours of sunshine and rainfall (millimeters) between 6AM and 11PM during the PA

[22] This reduces the number of child-day-school-time observations by 1.89%. The results are qualitatively robust to this exclusion.

[23] This reduces the number of child-day-leisure-time observations by 11.58%. The results are qualitatively robust to this exclusion.

measurement week. Rainfall is transformed into an indicator variable with the value 1 if there was more than 0 millimeters rainfall. The weather indicators are based on information from the weather station from the Royal Dutch Meteorological Institute (KNMI) located at Maastricht-Aachen airport and calculated for every day of the measurement week.

D. Summary statistics

Table A4 in the Appendix lists descriptive statistics of our full sample, i.e. of children in the target group who are enrolled in one of the 21 Active Living schools, as well as of their peers in non-Active Living schools. Firstly, the table shows that the children in Active Living schools account for 17 percent of our full sample. It also highlights that Active Living schools are located in low-SES areas; average school performance is significantly lower in Active Living schools as compared to the other (non-Active Living) schools in our data. Children in Active Living and non-Active Living schools are comparable with respect to age, gender, and grade in which they were enrolled in 2012/2013.

Table A5 in the Appendix lists descriptive statistics for the estimation samples of our main results (i.e., including child fixed-effects and a vector of control variables as specified in Section II). In our estimation samples, we have school performance data from 1,014 children, school time PA data from 536 children, and leisure time PA data from 509 children. When the school performance data are merged with the PA data, school time PA data remain for 422 children, and leisure time PA data for 398 children.

In the school performance data, 52 percent of the sample is enrolled in a treatment school. This share is slightly larger in the PA data (56 percent), which indicates that there are relatively more children in control schools for whom we do not have (valid) PA data.

Children on average took five tests per year (i.e., the number of tests on which the overall school performance measure is based). The average overall school performance in the pre-treatment year is 57.28, with a standard deviation of 7.90. Since the subject-scores are standardized across school years and children's performance generally improves as they get older, overall scores increase over time, to an average of 65.16 with a standard deviation of 8.04 in the post-treatment year.

The children in the estimation sample were on average 10.5 years old (126 months) when they took their pre-treatment tests, and 11.5 years old (138 months) when they took tests in the post-treatment year. Because we exclude the tests that were taken during the PA measurement weeks, the age of the children in the school performance data is on average 2 months lower than in the PA data.

In each data set, a little less than half the sample is male. The slight over-representation of the 5th graders cohort in the school performance data shows that there are more 4th graders than 5th graders who did not take any tests.

Children on average wore their accelerometer for four days during each PA measurement week. During these days, they wore their accelerometer for approximately five hours per day during school hours (309 minutes) in the pre-treatment year, out of which they spent one hour and a half (93 minutes) on PA. During leisure time in the pre-treatment year, children wore their accelerometer for approximately nine hours per day (548 minutes), out of which they spent three and one-half hours (204 minutes) on PA. Between the pre and post-treatment year, average time spent on PA during school time did not change significantly, but average leisure time PA decreased by 21 minutes per day.[24]

[24] Note that this trend is controlled for in the difference-in-differences estimations.

The weather conditions were slightly better during the PA measurements in the treatment year as compared to the PA measurements in the baseline year: outside temperatures were slightly higher and there were more hours of sunshine.

Figure A2 and Figure A3 in the Appendix indicate that the pre-treatment measures of school performance and time spent on PA are approximately normally distributed.

IV. Results

We firstly show the naive OLS estimates of the relationship between time spent on PA and overall school performance before the Active Living interventions were implemented, i.e., in 2012/2013. The results in Table 2 show that, in our sample, every additional 10 minutes spent on PA during school time is related to 0.52 points lower test scores (5.6 percent of a standard deviation in the full sample). This negative correlation between school time PA and school performance is significantly different from zero at the 1% level. Pre-treatment leisure time PA is not significantly related to pre-treatment school performance.

[Insert Table 2 Here]

However, these analyses could be biased due to reverse causality or omitted variables. It is possible that children who spend less time on PA at school use that time to invest in other cognitively stimulating activities such as studying, reading, or playing music. To account for this, we will now turn to our difference-in-differences estimations based on the Active Living program (see Table 3).

The difference-in-differences estimations reveal that Active Living has a negative effect on school performance, which is significantly different from zero at the 5% level. Due to Active Living, the increase in overall school performance of children in the treatment schools is 1.17

points smaller (0.13 standard deviations), as compared to the increase of those in the control schools. This result is robust across all specifications, i.e. without control variables, including child fixed-effects, controlling for age as well as child fixed-effects, and when we adjust p-values for the small number of clusters.

[Insert Table 3 Here]

Because of the larger variety of language tests as compared to math tests in our data, it is possible that our results are mostly driven by the performance on language tests. We therefore next conduct the analyses separately for language and math tests. Table A6 in the Appendix shows that the effect of Active Living is similar for school performance on language tests (-0.12 standard deviations) and math tests (-0.15 standard deviations). The results with respect to math tests should, however, be interpreted with caution, since we only find weak support for the parallel trends assumption for school performance on math tests (see Table A7 in the Appendix).[25]

Next, we analyze whether the treatment effects are heterogeneous across the school performance distribution, i.e., we investigate whether the Active Living effect on school performance is different for the worst-performing students than for the best-performing students. We estimate these heterogeneous effects nonlinearly, using a moving window of 200 individuals across the overall school performance distribution in the pre-treatment year 2012/2013.[26]

[Insert Figure 3 Here]

[25] The pre-treatment trends in school performance on language tests are not significantly different across treatment and control schools (see Table A7 in the Appendix), which suggests that we can indeed use a difference-in-differences technique to estimate the effects on language performance.

[26] Again, we first verified whether the parallel trends assumption holds for each selection of 200 individuals. Figure A4 in the Appendix shows that this is the case across the vast majority of the distribution: only in the right tail of the school performance distribution do the differences between the control and treatment group change significantly between the pre-treatment years.

For each selection of 200 individuals, the estimated Active Living effect on school performance is depicted on the y-axis of Figure 3. The Figure indicates that Active Living negatively affects school performance across the distribution. However, the negative effect is strongest in the left tail of the distribution, indicating that among those who were performing worst at school during the pre-treatment year, Active Living had the strongest negative effect.

A. Robustness checks

***Checking for outliers*.—**Firstly, we verify whether our results may be driven by an outlier school. This is of particular concern due to the relatively small number of schools included in this study, and because the treatment is not uniform across the treatment schools. Figure 4 provides indications that it is unlikely that the results are driven by any particular treatment school, i.e. any particular intervention package.[27] We next run regressions excluding each time one different school (see Tables A8 and A9 in the Appendix). Considering that the results remain robust across these regressions, we conclude that outlier treatment or control schools do not drive the estimates.

[Insert Figure 4 Here]

***Placebo interventions*.—**To investigate whether the effects that are found are indeed due to the Active Living interventions, we assign placebo interventions to the pre-treatment years 2012/2013, 2011/2012 and 2010/2011.[28] For each placebo treatment, we compare the test scores of the Active Living target group in two consecutive years. The model we estimate is as follows:

(3) $Score_{ist} = \delta_0 + \delta_1 DiD_s^* + \delta_2 Treat_s + \delta_3 Post_t^* + \delta_4 age_{ist} + \delta_5 age^2{}_{ist} + \gamma_{3is} + \varepsilon_{3ist}$

[27] It is not possible to identify the effects of one particular intervention, since each treatment school has its own specific intervention package. Identifying the effects of each intervention package separately (i.e., including each time one different treatment school) indicates that the treatment effect on school performance is never positive (results available upon request). However, this strategy reduces the number of observations drastically, which makes the interpretation of insignificant effects problematic due to a lack of power.

[28] Note that the actual interventions were implemented between June 2013 and March 2014.

*Post** is an indicator variable with value 1 in the post-placebo-treatment year and 0 in the pre-placebo-treatment year. The *DiD** variable is the interaction between *Post** and *Treat*. Since no actual interventions were implemented in the pre-treatment years, we expect to find that we cannot reject the hypothesis that the difference-in-differences estimator δ_1 is equal to zero.

[Insert Table 4 Here]

The results in Table 4 show that, as expected, Active Living does not have significant effects in any of the pre-treatment years. This is a strong indication that it is indeed the implementation of the Active Living interventions which is driving the effect on school performance.

***Comparisons to non-Active Living schools*.—**Next, we analyze the change in the difference with respect to school performance between children in Active Living schools and children in schools which did not participate in the Active Living program as a treatment or control group. By doing so, we verify that the effects are indeed coming from a change in school performance of the treatment group, and not from a (unexpected) change in school performance of the control group. Non-Active Living schools are all schools included in our data that did not participate in the Active Living study as a treatment or control group.[29]

When we compare the school performance of children in the treatment group to those of children in non-Active Living schools, the difference-in-differences model we estimate is similar to equation (1), with the exception that the *Treat* and *DiD* variables now have the value 0 for children in the non-Active Living schools and are set to missing for children in the control schools.

When we compare the children in the control schools to those in the non-Active Living schools, the difference-in-differences model we estimate is as follows:

[29] Before we run this analysis, we verified that the pre-treatment trends with respect to test scores are indeed similar between both groups (see Table A10 in the Appendix).

(4) $Score_{ist} = \delta_6 + \delta_7 DiD'_s + \delta_8 Control_s + \delta_9 Post_t + \delta_{10} age_{ist} + \delta_{11} age^2{}_{ist} + \gamma_{4is} + \varepsilon_{4ist}$

Control is an indicator variable with value 1 for children in control schools, and 0 for children in non-Active Living schools. The *DiD'* variable is the interaction between *Control* and *Post*. The other variables are identical to those in equation (1).

Although pre-treatment differences in school performance exist because Active Living schools are located in relatively deprived areas (as was shown in Table A4 in the Appendix), these differences should not change significantly over time. Any significant effects that are found based on this analysis indicate that the control group may be (unexpectedly) affected by Active Living, which would complicate the interpretation of any Active Living effects we find. However, if we indeed cannot reject the hypothesis that the difference-in-differences estimator δ_7 is equal to zero, we have strong reasons to believe that the Active Living effects are the result of the Active Living interventions in the treatment schools. This conclusion becomes even stronger if the point estimate of the Active Living effect is similar to our main result (i.e., negative) when we compare the treatment group to the non-Active Living schools.

The results in Table 5 show that the school performance of children in the control group slightly increases due to the Active Living program (0.07 standard deviations), but as expected, this effect is not significantly different from zero. The effect of Active Living on the treatment group's school performance remains negative and significantly different from zero at the 10% level in the main specification, although it is now smaller in size (-0.06 standard deviations).

[Insert Table 5 Here]

This smaller effect may be explained by the positive, though insignificant, Active Living effect on the control group's school performance. Although caution is needed when interpreting results that are not statistically significant, this suggests that the estimated effects as presented in Table

3 might be slightly biased upwards, and that the more conservative estimate of the average effect of Active Living on overall school performance is 6 percent of a standard deviation.[30]

B. Mechanism

The robustness checks give strong reasons to believe that the negative effect of Active Living on school performance is due to the Active Living interventions in the treatment schools. Considering that the Active Living interventions aim to decrease sedentary behavior, we hypothesize that time spent on PA is an important mechanism for this effect. This also means that, in the setting we study, it may still be possible to conclude that PA has a positive effect on school performance, if Active Living (unexpectedly) has a negative effect on time spent on PA. In order to analyze this, we make use of the accelerometer data.[31]

Table 6 reports the results of the difference-in-differences analyses on time spent on PA during school time and leisure time. Based on these results, we conclude that Active Living significantly increases time spent on PA during school hours by 9.3 minutes per school day. This effect is equal to 0.34 standard deviations. There appears to be a small crowding-out effect of leisure time PA, as Active Living decreases time spent on PA during leisure time by 3.8 minutes per day (0.08 standard deviations), but this effect is not significantly different from zero.

[Insert Table 6 Here]

Figure A5 in the Appendix shows the estimated effects of Active Living on time spent on PA during school time, across the pre-treatment overall school performance distribution. Again, the estimated effects are plotted on the y-axis for each selection of 200 individuals. The figure indicates that Active Living has a positive effect on time spent on PA during school hours of

[30] We also ran a regression of school performance on participating in the Active Living program as a treatment school in which we control for baseline school performance instead of including the *DiD* variable. The estimated treatment effect using this specification is -0.06 standard deviations, significantly different from zero at the 10 percent level. Results are available upon request.

[31] Additional analyses (available upon request) indicate that wearing an accelerometer does not, in itself, have any statistically or economically significant effects on school performance.

approximately 10 minutes per day across the school performance distribution. Active Living does not have a significant effect on leisure time PA in almost all parts of the school performance distribution (see Figure A6 in the Appendix).

These results can help to interpret the size of the effect of PA on school performance. Under the assumption that Active Living only affects school performance via school time PA, we would conclude that an extra 30 minutes spent on PA during school hours reduces overall school performance by 19 percent of a standard deviation.[32] This assumption may appear to be strong, since it is conceivable that Active Living also impacted other potential confounders (e.g., increased leisure time PA, and improved health and well-being from adjusted playgrounds or safer roads around school). However, as these examples indicate, it is implausible that such confounders are negatively related to school performance. They can therefore not explain the negative effect on school performance; controlling for such variables would make the negative effect on school performance even stronger. Considering the robust negative effects of Active Living on school performance we find, we therefore conclude that it is most likely that Active Living crowds-out time investments in (other) cognitively stimulating activities during school hours such as focusing and paying attention during instruction hours.

V. Conclusions

This paper investigates whether stimulating physical activity in everyday life affects primary school performance. Our estimations are based on the *Active Living* field experiment, which aims to increase physical activity and decrease sedentary behavior among 8-12 year-olds living in

[32] Using results from Table 5 (column 6) and Table 6 (column 3), a back of the envelope calculation of the Wald estimator is -0.059(-0.546/9.298), implying that for each minute increase in school time PA, overall school performance decreases by 0.059 points (0.63 percent of a standard deviation). Note that such an interpretation requires the instrument to be strong. Although the effect of Active Living on school time PA is significant (t=2.89), it does not meet the criterion of a strong instrument.

low-SES areas in the Netherlands. The interventions focus on active transportation and active play at school and in leisure time. Results from difference-in-differences analyses reveal that Active Living significantly decreases overall school performance by 6 percent of a standard deviation. This result is robust across a variety of specifications, but the negative effect is strongest among the worst-performing students. The results of several robustness checks provide strong indications that the negative effect is indeed due to the Active Living interventions causing a significant change in the treatment group's school performance.

Besides its effects on school performance, Active Living causes a significant increase in time spent on physical activity during school time of 9.3 minutes per school day (0.34 standard deviations). To put this into perspective: the average amount of time spent on physical activity during school hours is 93 minutes per day. The effect remains robust across the pre-treatment school performance distribution. Leisure time physical activity is not significantly affected by Active Living. We conclude that increased time spent on physical activity during school hours is an important mechanism for the negative effect on school performance we find.

Our results shed new light on the effects of stimulating physical activity among children and adolescents. To our knowledge, this study is the first to analyze the causal effect of stimulating informal physical activity on educational outcomes. Previous studies have shown a positive relationship between formal physical activity (i.e., doing sports, or physical education) and school performance, but based on our findings, we conclude that these results appear not to be generalizable to informal physical activity such as active transport or active play. One reason for this may be that formal physical activity is more effective for fostering skills that are beneficial for school performance, such as confidence, discipline or competitiveness, than informal physical activity. Moreover, because exercising is generally of higher intensity physical activity than being active in everyday life, formal physical activity may yield higher returns in terms of health and well-being. If children usually reallocate study hours (or other cognitively stimulating tasks such as paying attention during instruction hours) to physical activity, the combination of the absence of the development of sports-related skills with the potentially lower health benefits, could explain why stimulating informal physical activity has a negative effect on school performance.

Additionally, our results indicate that stimulating informal physical activity may increase inequality in school performance. An important characteristic of Active Living is that it focuses on children living in low-SES areas, where (pre-treatment) school performance is lower. This means that Active Living negatively affects school performance in the worse performing regions of the Netherlands, and that, within these regions, the negative effect is strongest among the worst-performing students.

Nevertheless, although we find strong evidence for negative effects of Active Living on school performance, it is a relatively short-term treatment effect (up to one year post-treatment). We cannot exclude the possibility that, in the longer run, Active Living will have positive effects on educational outcomes.

Taken together, our study reveals an important caveat for researchers and policy makers in the fields of health and education. Research and policy which focus solely on the effects of interventions on physical activity may overlook the negative externalities these interventions may have. Increased time spent on physical activity decreases time left available for other (desirable) activities. In the case of Active Living, we see that although it succeeded in increasing time spent on physical activity, it came at the cost of decreased school performance. Before resources are allocated to increase time spent on physical activity among children and adolescents, we need to improve our understanding of the causal effects of such interventions on educational outcomes. Future research needs to focus on which forms of physical activity do not crowd-out desirable activities.

References

Angrist, J.D., and Pischke, J.S. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Barron, J., Ewing, B., and Waddell, G. 2000. "The effects of high school athletic participation on education and labor market outcomes." *Review of Economics and Statistics* 82(3): 409-421.

Bell, R.M., and McCaffrey, D.F. 2002. "Bias reduction in standard errors for linear regression with multi-stage samples." *Survey Methodology* 28(2): 169-181.

Berg, V. van den, Salasi, E., Groot, R.H.M. de, Jolles, J., Chinapaw, M.J.M., and Singh, A.S. 2016. "Physical activity in the school setting: Cognitive performance is not affected by three different types of acute exercise." *Frontiers in Psychology* 7(723): 1-9.

Bertrand, M., Duflo, E., and Mullainathan, S. 2004. "How much should we trust difference-in-differences estimates?" *The Quarterly Journal of Economics* 119(1): 249-275.

Cabane, C., Hille, A., and Lechner, M. 2016. "Mozart or Pelé? The effects of adolescents' participation in music and sports." *Labour Economics* 41: 90-103.

Cameron, A.C., Gelbach, J.B., and Miller, D.L. 2008. "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics* 90(3): 414-427.

Cameron, A.C. and Miller, D.L. 2015. "A practitioner's guide to cluster-robust inference." *The Journal of Human Resources* 50(2): 318-372.

Cawley, J., Frisvold, D., and Meyerhoefer, C. 2013. "The impact of physical education on obesity among elementary school children." *Journal of Health Economics* 32(4): 743–755.

Cawley, J., Meyerhoefer, C., and Newhouse, D. 2007. “The impact of state physical education requirements on youth physical activity and overweight.” *Health Economics* 16(12): 1287–1301.

Centers for Disease Control and Prevention. 2011. *Strategies to prevent obesity and other chronic diseases: the CDC guide to strategies to increase physical activity in the community*. Atlanta, U.S.A.: U.S. Department of Health and Human Services.

Choi, L., Liu, Z., Matthews, C.W., and Buchowski, M.S. 2011. “Validation of accelerometer wear and nonwear time classification algorithm.” *Medicine, and Science in Sports, and Exercise* 43(2): 357-364.

Colcolmbe, S. and Kramer, A.F. 2003. “Fitness effects on the cognitive functioning of older adults: A meta-analytic study.” *Psychological Science* 14(2): 125-130.

Dills, A.K., Morgan, H.N., and Rotthoff, K.W. 2011. “Recess, physical education, and elementary school student outcomes.” *Economics of Education Review* 30: 889-900.

Donald, D.G., and Lang, K. 2007. “Inference with difference-in-differences and other panel data.” *The Review of Economics and Statistics* 89(2): 221-223.

Eide, E. and Ronan, N. 2001. “Is participation in high school athletics an investment or a consumption good? Evidence from high school and beyond.” *Economics of Education Review* 20(5): 431-442.

Erickson, K. I., Hillman, C. H., and Kramer, A. F. 2015. “Physical activity, brain, and cognition.” *Current Opinion in Behavioral Sciences* 4: 27-32.

European Commission. 2014. *EU action plan on childhood obesity 2014-2020.*

Evenson, K., Catellier, D., Gill, K., Ondrak, K., and McMurray, R. 2008. “Calibration of two objective measures of physical activity for children.” *Journal of Sports Sciences* 26: 1557–1565.

Felez-Nobrega, M., Hillman, C.H., Cirera, E., and Puig-Ribera, A. 2017. “The association of context-specific sitting time and physical activity intensity to working memory capacity and academic achievement in young adults.” *The European Journal of Public Health* 27(4): 741-746.

Felfe, C., Lechner, M., and Steinmayr, A. 2016. “Sports and child development.” *PLoS ONE* 11(5): 1-23.

Hillman, C.H., Erickson, K.I., and Kramer, A.F. 2008. “Be smart, exercise your heart: exercise effects on brain and cognition.” *Nature Reviews Neuroscience* 9: 58-65.

Kann, D.H.H. van, Jansen, M.W.J., Vries, S.I. de, and Kremers, S.P.J. 2015. “Active Living: development and quasi-experimental evaluation of a school-centered physical activity intervention for primary school children.” *BMC Public Health* 15(1315): 1-10.

Kann, D.H.H. van, Kremers, S.P.J. Jansen, Vries, N.K. de, Vries, S.I. de, and Jansen, M.W.J. 2016. “The effect of a school-centered multicomponent intervention on daily physical activity and sedentary behavior in primary school children: the Active Living study.” *Preventive Medicine* 89: 64-69.

Lechner, M. 2009. “Long-run labour market and health effects of individual sports activities.” *The Journal of Health Economics* 28: 839–854.

Lipscomb, S. 2007. “Secondary school extracurricular involvement and academic achievement: a fixed effects approach.” *Economics of Education Review* 26: 463-472.

Lizandra, J., Devís-Devís, J., Pérez-Gimeno, E., Valencia-Peris, A., and Peiró-Velert, C. 2016. “Does sedentary behavior predict academic performance in adolescents or the other way round? A longitudinal path analysis.” *PLoS ONE* 11(14): 1-13.

Pfeifer, C. and Cornelissen, T. 2010. “The impact of participation in sports on educational attainment – new evidence from Germany.” *Economics of Education Review* 29(1): 94-103.

Rees, D.I., and Sabia, J.J. 2010. "Sports participation and academic performance: Evidence from the National Longitudinal Study of Adolescent Health." *Economics of Education Review* 29: 751-759.

Rich, C., Griffiths, L., and Dezateux, C. 2012. "Seasonal variation in accelerometer-determined sedentary behaviour and physical activity in children: a review." *International Journal of Behavioral Nutrition and Physical Activity* 9(49): 1-8.

Singh, A., Uijtdewilligen, L., Twisk, J.W.R., Mechelen, W. van, and Chinapaw, M.J.M. 2012. "Physical activity and performance at school: A systematic review of the literature including a methodological quality assessment." *Archives of Pediatrics and Adolescent Medicine* 166(1): 49-55.

Statistics Netherlands. 2013. Statistics on population density via statline.cbs.nl.

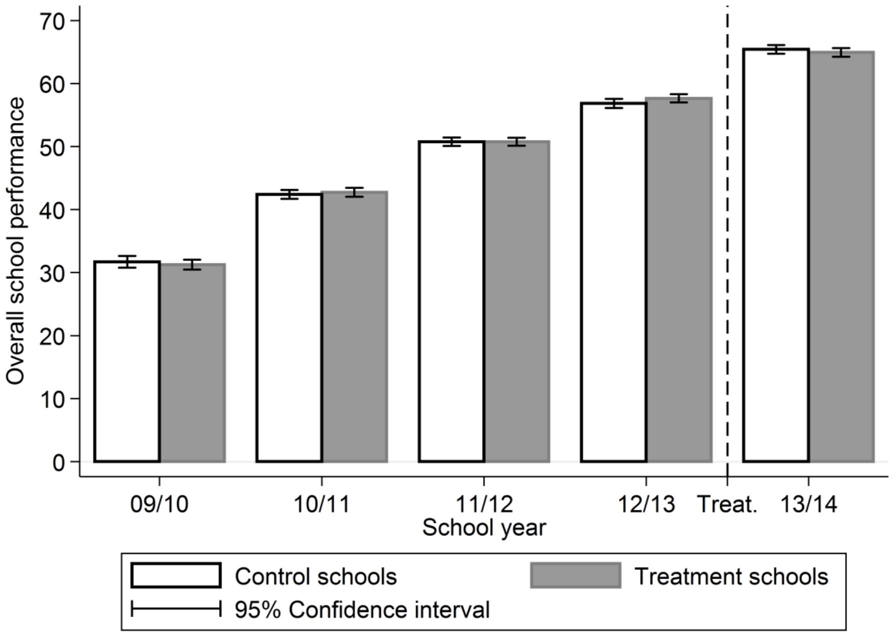
Statistics Netherlands. 2014. Statline: nabijheid voorzieningen; afstand locatie, wijk- en buurtcijfers 2013. Voorburg/Heerlen: Statistics Netherlands.

Stevenson, B. 2010. "Beyond the classroom: using Title IX to measure the return to high school sports." *Review of Economics and Statistics* 92(2): 284-301.

Strong, W., Malina, R., Blimkie, C., Daniels, S., Dishman, R., and Gutin, B. 2005. "Evidence based physical activity for school-age youth." *The Journal of Pediatrics* 146(6): 732-737.

World Health Organization. 2010. *Global recommendations on physical activity for health.* Geneva, Switzerland: WHO press.

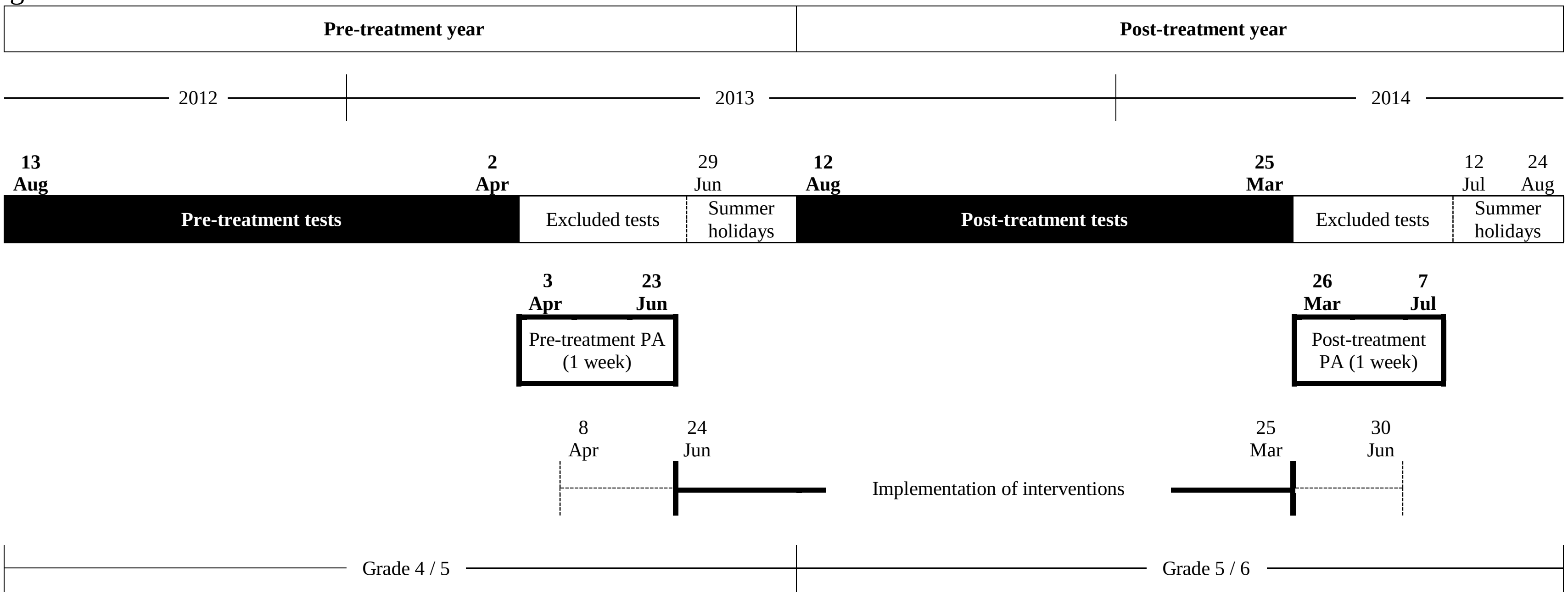
Figure 1. Trends in overall school performance in treatment and control schools



Notes: Tests taken between April 3rd and August 11th 2013 are excluded from the analyses. The results in Table A3 (see Appendix) show that the difference between treatment and control schools does not change significantly in the pre-treatment years.

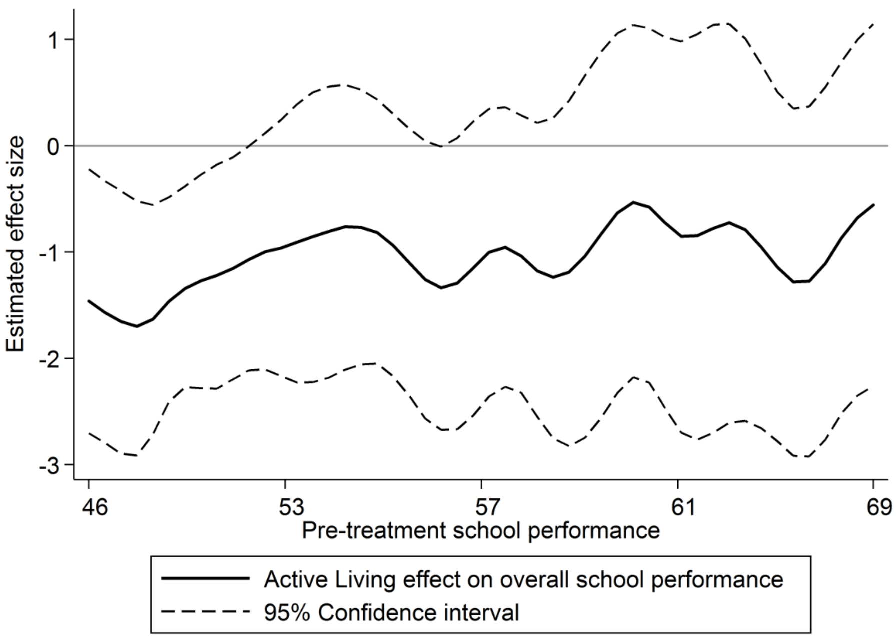
Source: School performance data.

Figure 2. Timeline of the data collection



Notes: In each school, PA data was collected during one week between April-June 2013 and again during one week between March-July 2014. Treatment schools could implement their interventions between the last day of the pre-treatment PA measurement week in 2013 and the first day of the post-treatment PA measurement week in 2014. Implementation periods can therefore differ by school as indicated by the dashed lines. In some analyses, pre-treatment test scores taken before August 13th 2012 are also included.

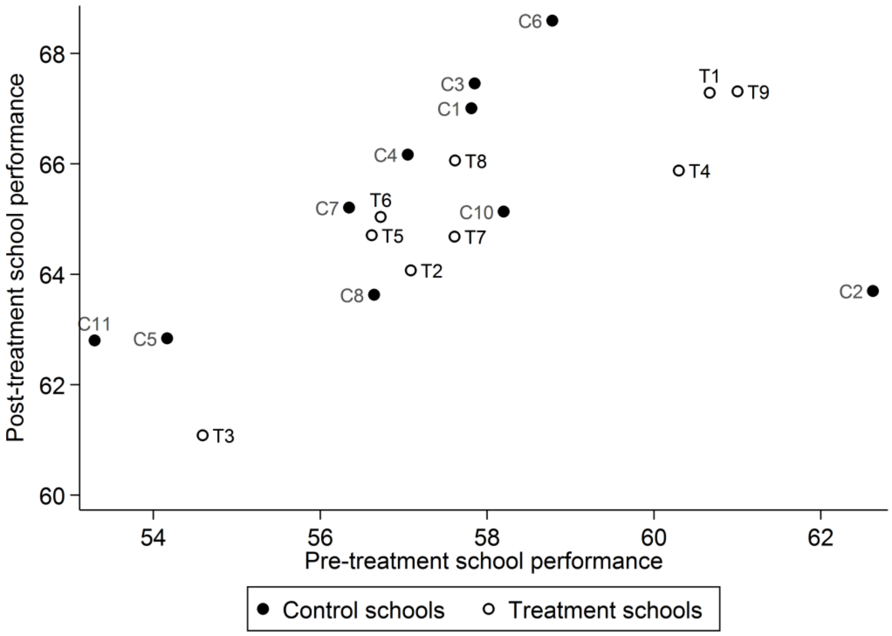
Figure 3. The effect of Active Living on overall school performance, by pre-treatment school performance



Notes: Local polynomial smooth graph of the results of ordinary least squares regression analyses across the pre-treatment overall school performance distribution in 2012/2013. Pre-treatment school performance is a moving window of 200 child IDs. The dependent variable in each regression is overall school performance. The estimated effect size based on difference-in-differences estimations is plotted on the y-axis. Independent variables are a year dummy, age, age squared and child fixed-effects. Adjusted 95% confidence intervals for estimated coefficients with robust standard errors are calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution, because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure A4 in the Appendix).

Source: School performance data.

Figure 4. Average pre and post-treatment overall school performance, by Active Living school



Notes: Based on the estimation sample of the analysis presented in Table 3, column (3). The numbers of the treatment schools correspond to those in Table 1. The results in Tables A8 and A9 (see Appendix) show that the difference-in-differences estimations of the Active Living effect on overall school performance remain robust when one of the Active Living schools is excluded from the analysis.

Source: School performance data.

Table 1. Overview of implemented interventions per treatment school

	T1	T2	T3	T4	T5	T6	T7	T8	T9
Panel A. Active transportation to school									
Development safe route to school	−	−	+	+	+	−	−	−	+
Mobilize crossing guards	−	−	+	+	−	−	−	−	−
Adapt unsafe intersection in school environment	−	−	+	−	−	−	−	−	−
Availability of bicycle racks	−	+	−	−	−	+	−	−	−
School pedestrian crossing indicators	+	+	−	−	−	+	−	−	−
Create safer parking situation around school	−	−	+	−	−	−	−	−	−
Create traffic circle in schoolyard environment	+	+	−	−	−	−	−	−	−
Sticker competition for active transport	+	+	+	+	−	−	+	−	−
School stimulation documentation on safe active transport	−	−	+	+	−	−	−	−	−
Introduction 'Walk/Bike-to-school-day'	−	−	−	−	−	+	−	−	−
Speed check action performed by children	−	−	+	+	−	−	−	−	−
Lessons to improve bicycle skills	−	−	−	−	+	−	−	−	+
Panel B. PA in school									
New fixed equipment in schoolyard	−	+	−	+	+	+	−	+	+
New loose equipment in schoolyard	−	+	+	−	+	+	+	−	+
Playground markings	−	+	−	−	−	+	+	−	−
Establish ball game area	+	+	−	−	−	+	−	+	−
Put a ball backstop besides railway	+	−	−	−	−	−	−	−	−
Sound equipment in schoolyard environment	−	+	−	−	−	−	−	−	−
Additional sports day in schoolyard	+	+	+	+	−	−	−	+	−
Sports clinics in recess	+	+	+	+	+	+	−	−	+
Use of schoolyard games	−	+	+	−	−	−	+	+	−
Prize contest for best idea PA stimulation children	+	−	−	−	−	−	−	+	−
Panel C. PA in leisure time									
Establish training circuit	−	−	+	−	−	−	−	−	−
Active Living Games	+	+	+	+	+	+	+	−	+
Establish out-of-school PA program	+	+	+	+	+	−	−	+	+
Establish school soccer team	−	−	+	−	−	−	−	−	−
Establish PA activities by children for local residents	−	−	+	−	−	−	−	−	−

Notes: Tn = Treatment school (number); + = intervention was implemented; − = intervention was not implemented. One of the initial ten treatment schools is excluded from the analyses because it organized a sports day during the baseline physical activity measurement week. Control schools did not implement any interventions. See Table A1 in the Appendix for a short explanation of each intervention.

Table 2. Pre-treatment correlation between time spent on physical activity and school performance

Dep. var. : overall school performance	(1)	(2)	(3)	(4)	(5)	(6)
	PA during school time			PA during leisure time		
	Base	Incl. controls from school perf. data	Incl. controls from PA data	Base	Incl. controls from school perf. data	Incl. controls from PA data
Time spent on PA (min./day)	-0.033	-0.031	-0.052	-0.008	-0.007	-0.009
	(0.013)	(0.011)	(0.012)	(0.005)	(0.005)	(0.005)
Adjusted p-value	[0.018]	[0.012]	[0.004]	[0.142]	[0.214]	[0.116]
Age when tests were taken (months)		0.966	0.969		1.112	0.938
		(0.783)	(0.704)		(0.778)	(0.595)
Age squared		-0.005	-0.005		-0.005	-0.004
		(0.003)	(0.003)		(0.003)	(0.002)
Cohort (1=5th graders, 0=4th graders)		10.183	10.096		10.080	9.657
		(0.819)	(0.879)		(0.957)	(1.032)
Gender (1=boy)		0.089	0.326		0.193	0.255
		(0.981)	(0.962)		(1.100)	(1.027)
Wearing time of accelerometer (min./day)			0.013			0.129
			(0.101)			(0.050)
Wearing time squared			-0.000			-0.000
			(0.000)			(0.000)
Outside temperature during PA measurement week (Celsius*10)			-0.054			-0.052
			(0.031)			(0.040)
Temperature squared			0.000			0.000
			(0.000)			(0.000)
Amount of sunshine during PA measurement week (hours/day)			1.125			0.591
			(0.530)			(0.659)
Sunshine squared			-0.158			-0.081
			(0.085)			(0.101)
Rain during PA measurement week (1=yes)			-0.562			-0.507
			(0.924)			(1.011)
Constant	60.654	9.015	9.258	59.427	-2.588	-26.163
	(1.393)	(47.984)	(49.228)	(1.046)	(48.199)	(43.406)
Observations	549	549	512	533	533	500
R-squared	0.012	0.270	0.288	0.003	0.263	0.281
Mean school performance in estimation sample	57.665	57.665	57.668	57.743	57.743	57.767
Standard deviation	(7.795)	(7.795)	(7.687)	(7.827)	(7.827)	(7.681)
Mean school performance in full sample	57.803	57.803	57.803	57.803	57.803	57.803
Standard deviation	(8.304)	(8.304)	(8.304)	(8.304)	(8.304)	(8.304)

Notes: Results of six ordinary least squares regression analyses. The dependent variable in each column is overall school performance in 2012/2013. Tests taken between April 3rd and August 11th 2013 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.

Source: Merged school performance and PA data.

Table 3. Difference-in-differences estimations of the effect of Active Living on overall school performance

Dep. var. : overall school performance	(1) OLS	(2) FE	(3) FE
Treatment * year	-1.289	-1.090	-1.170
	(0.567)	(0.543)	(0.544)
Adjusted p-value	[0.037]	[0.061]	[0.048]
Treatment (1=Treatment group)	0.809		
	(0.846)		
Year (1=post-treatment year, 0=pre-treatment year)	8.589	8.448	2.950
	(0.471)	(0.444)	(3.798)
Age when tests were taken (months)			0.972
			(0.400)
Age squared			-0.002
			(0.001)
Child fixed-effects	No	Yes	Yes
Constant	56.860		
	(0.531)		
Observations	2,064	2,028	2,028
R-squared	0.199	0.809	0.812
Number of child ID		1,014	1,014
Mean school performance in estimation sample	61.312	61.222	61.222
Standard deviation	(8.910)	(8.890)	(8.910)
Mean school performance in full sample	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)

Notes: Results of three ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Tests taken between April 3rd and August 11th 2013 and between March 26th and August 14th 2014 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.

Source: School performance data.

Table 4. Placebo difference-in-differences estimations of the effect of Active Living on overall school performance

Dep. var. : overall school performance	(1)	(2)	(3)	(4)	(5)	(6)
	Placebo treatment in...					
	2012/2013		2011/2012		2010/2011	
	OLS	FE	OLS	FE	OLS	FE
Treatment * year	0.812	0.680	-0.339	-0.431	0.794	0.367
	(0.555)	(0.573)	(0.421)	(0.308)	(0.885)	(0.842)
Adjusted p-value	[0.164]	[0.253]	[0.433]	[0.182]	[0.384]	[0.669]
Treatment (1=Treatment group)	-0.003		0.336		-0.458	
	(0.874)		(0.859)		(1.094)	
Year (1=2012/2013, 0=2011/2012)	6.079	4.753				
	(0.414)	(3.727)				
Year (1=2011/2012, 0=2010/2011)			8.358	0.345		
			(0.294)	(3.027)		
Year (1=2010/2011, 0=2009/2010)					10.688	7.382
					(0.692)	(9.173)
Age when tests were taken (months)		0.067		1.814		2.656
		(0.656)		(0.383)		(0.948)
Age squared		0.000		-0.005		-0.012
		(0.002)		(0.001)		(0.001)
Child fixed-effects	No	Yes	No	Yes	No	Yes
Constant	50.781		42.422		31.735	
	(0.624)		(0.646)		(0.737)	
Observations	2,009	1,988	1,952	1,912	1,895	1,872
R-squared	0.153	0.779	0.222	0.836	0.291	0.850
Number of child ID		994		956		936
Mean school performance in estimation sample	54.061	45.787	46.770	45.636	37.102	45.613
Standard deviation	(8.349)	(12.625)	(8.685)	(12.659)	(10.312)	(12.688)
Mean school performance in full sample	54.578	54.578	47.198	47.198	37.624	37.624
Standard deviation	(8.769)	(8.769)	(9.259)	(9.259)	(10.548)	(10.548)

Notes: Results of six ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Tests taken between April 3rd and August 11th 2013 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.

Source: School performance data.

Table 5. Difference-in-differences estimations of the effect of Active Living on overall school performance – Children in Active Living schools compared to children in non-Active Living schools

Dep. var. : overall school performance	(1)	(2)	(3)	(4)	(5)	(6)
	Control group			Treatment group		
	OLS	FE	FE	OLS	FE	FE
Control * year	0.342	0.765	0.664			
	(0.501)	(0.453)	(0.466)			
Adjusted p-value	[0.496]	[0.094]	[0.157]			
Treatment * year				-0.090	-0.325	-0.546
				(0.880)	(0.334)	(0.301)
Adjusted p-value				[0.919]	[0.331]	[0.072]
Control (1=Control group, 0=non-Active Living)	-1.056					
	(0.606)					
Treatment (1=Treatment group, 0=non-Active Living)				-0.246		
				(0.715)		
Year (1=post-treatment year, 0=pre-treatment year)	8.247	7.683	-1.351	8.247	7.683	-1.080
	(0.199)	(0.133)	(1.465)	(0.199)	(0.133)	(1.487)
Age when tests were taken (months)			1.566			1.571
			(0.190)			(0.185)
Age squared			-0.003			-0.003
			(0.001)			(0.001)
Child fixed-effects	No	Yes	Yes	No	Yes	Yes
Constant	57.916			57.916		
	(0.313)			(0.313)		
Observations	11,361	10,350	10,348	11,480	10,438	10,436
R-squared	0.197	0.788	0.801	0.195	0.784	0.797
Number of child ID		5,175	5,174		5,219	5,218
Mean school performance in estimation sample	62.311	61.665	61.667	62.367	61.687	61.689
Standard deviation	(9.313)	(9.173)	(9.173)	(9.309)	(9.143)	(9.143)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: Results of six ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Tests taken between April 3rd and August 11th 2013 and between March 26th and August 14th 2014 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.

Source: School performance data.

Table 6. Difference-in-differences estimations of the effect of Active Living on time spent on physical activity during school time and leisure time

Dep. var. : time spent on PA (min./day)	(1)	(2)	(3)	(4)	(5)	(6)
		School time			Leisure time	
	OLS	FE	FE	OLS	FE	FE
Treatment * year	25.337	22.544	9.298	-2.247	-5.638	-3.814
	(8.552)	(8.611)	(3.283)	(7.836)	(7.126)	(3.853)
Adjusted p-value	[0.009]	[0.017]	[0.016]	[0.778]	[0.439]	[0.344]
Treatment (1=Treatment group)	-14.990			1.414		
	(5.243)			(5.293)		
Year (1=post-treatment, 0=pre-treatment)	-15.175	-13.062	-13.963	-18.765	-19.374	-24.136
	(5.872)	(6.179)	(3.182)	(4.426)	(4.494)	(5.104)
Wearing time of accelerometer (min./day)			0.790			0.412
			(0.305)			(0.191)
Wearing time squared			-0.001			-0.000
			(0.000)			(0.000)
Outside temperature during PA measurement week (Celsius*10)			-0.269			0.184
			(0.069)			(0.146)
Temperature squared			0.001			-0.000
			(0.000)			(0.001)
Amount of sunshine during PA measurement week (hours/day)			8.483			10.339
			(2.258)			(4.660)
Sunshine squared			-0.771			-0.853
			(0.268)			(0.520)
Rain during PA measurement week (1=yes)			1.441			-2.597
			(1.820)			(4.654)
Child fixed-effects	No	Yes	Yes	No	Yes	Yes
Constant	100.280			200.506		
	(4.111)			(3.049)		
Observations	1,355	1,204	1,072	1,301	1,120	1,018
R-squared	0.058	0.111	0.503	0.041	0.137	0.435
Number of child ID		602	536		560	509
Mean time spent on PA in estimation sample	91.510	91.973	93.500	191.819	192.564	192.978
Standard deviation	(26.989)	(26.756)	(25.183)	(49.545)	(48.696)	(46.767)
Mean time spent on PA in full sample	91.510	91.510	91.510	191.819	191.819	191.819
Standard deviation	(26.989)	(26.989)	(26.989)	(49.545)	(49.545)	(49.545)

Notes: Results of six ordinary least squares regression analyses. The dependent variable in columns (1)-(3) is time spent on PA during school time (min./day). The dependent variable in columns (4)-(6) is time spent on PA during leisure time (min./day). Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.

Source: PA data.

Appendices

For online publication only

Figure A1. Active Living research area – Southern-Limburg region, The Netherlands

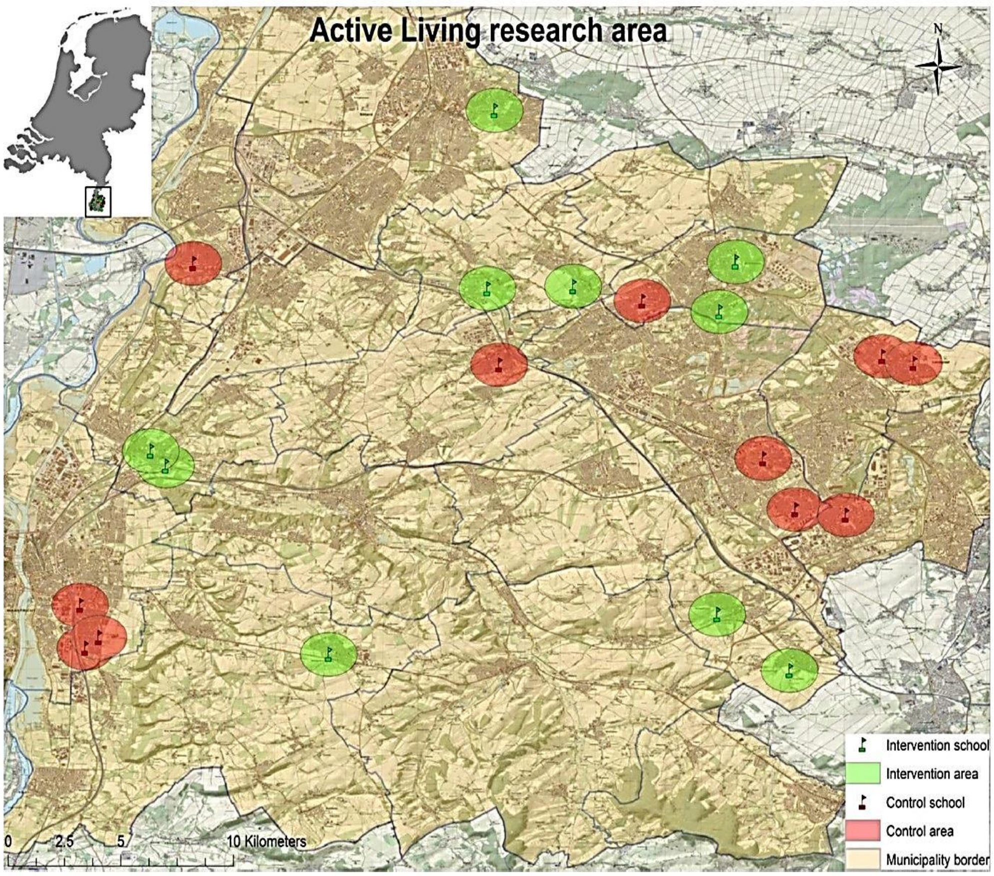
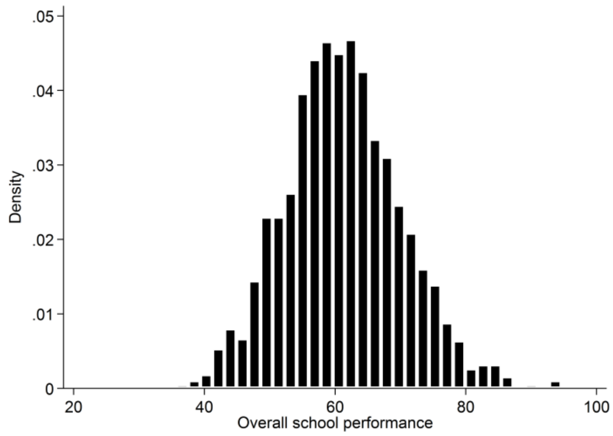
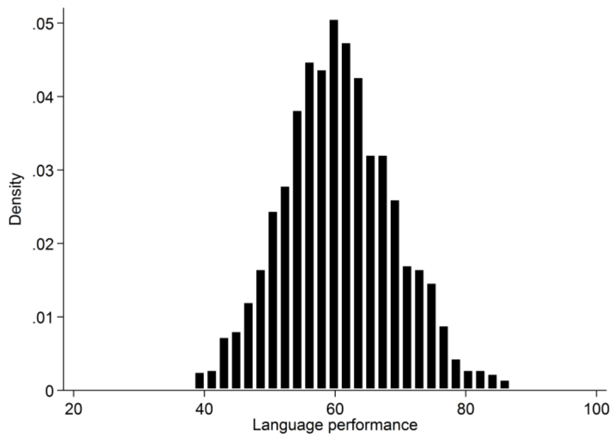


Figure A2. Distribution of school performance in the estimation sample

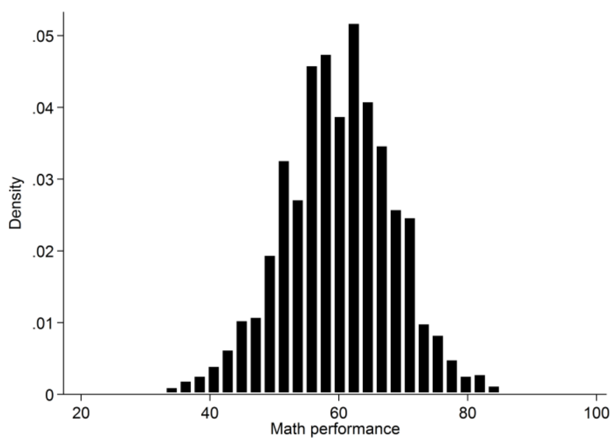
Panel A. Overall school performance



Panel B. Language



Panel C. Math

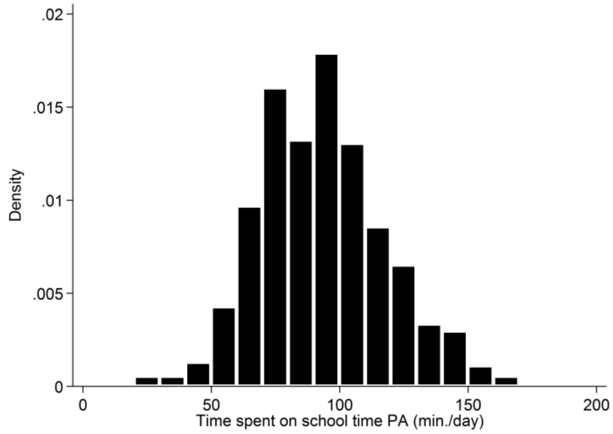


Notes: Based on the estimation sample of the analysis presented in column 3 of Table 3.

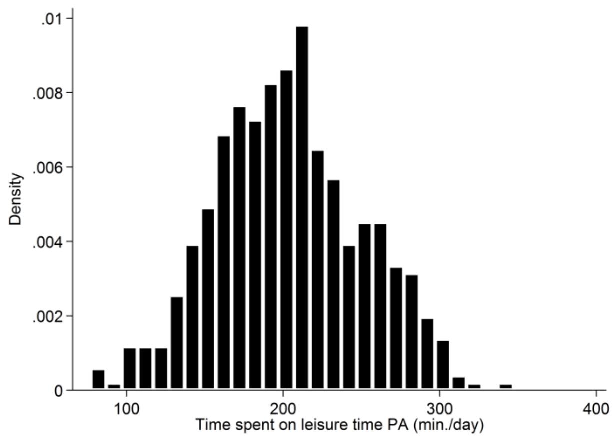
Source: School performance data.

Figure A3. Distribution of time spent on physical activity in the estimation sample

Panel A. School time PA



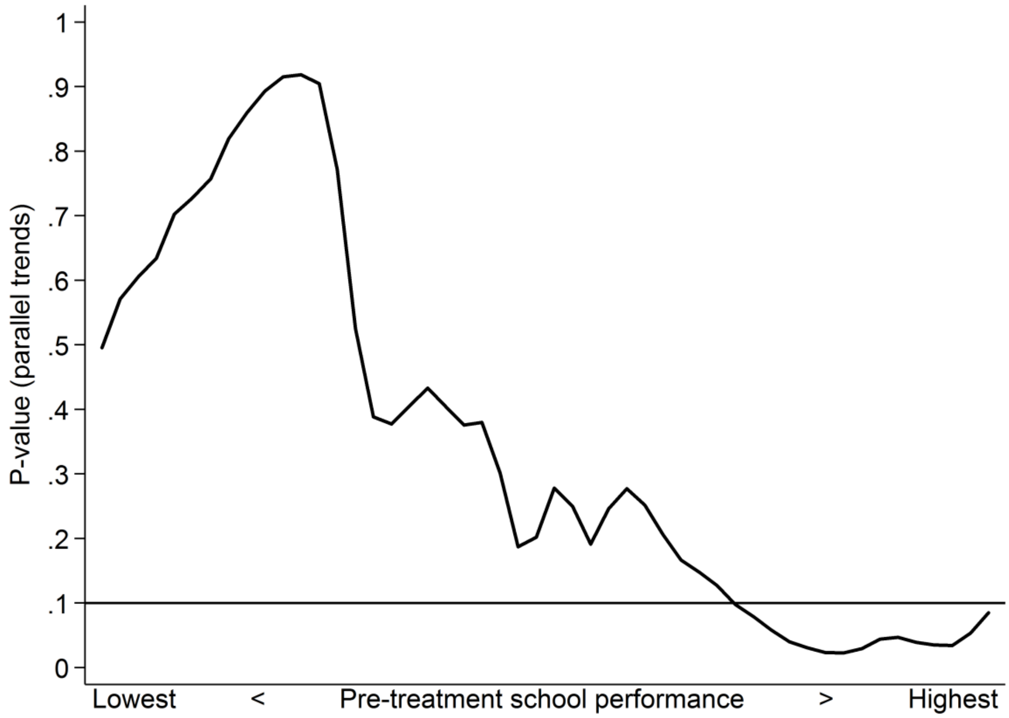
Panel B. Leisure time PA



Notes: Based on the estimation sample of the analyses presented in columns 3 and 6 of Table 6.

Source: PA data.

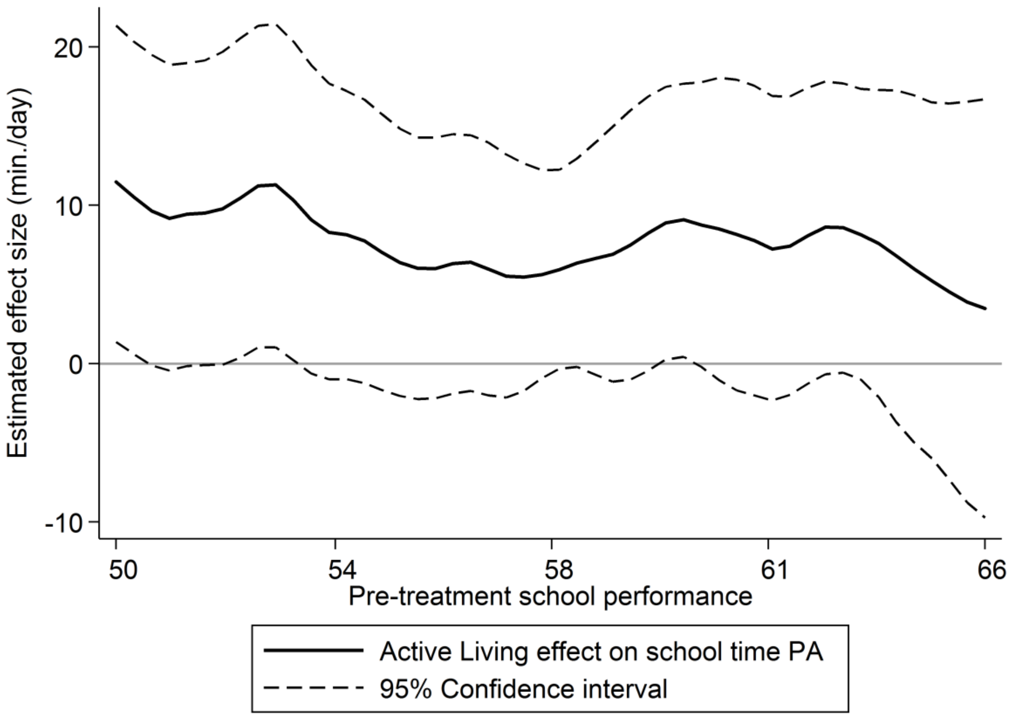
Figure A4. Tests of the parallel trends assumption, by pre-treatment school performance (2012/2013)



Notes: Local polynomial smooth graph of p-values for F-tests for joint significance of Treatment*year interactions. P-values are obtained after ordinary least squares regression analyses across the overall school performance distribution in the pre-treatment year (2012/2013). Pre-treatment school performance is a moving window of 200 child IDs. Independent variables are year dummies for the 2009/2010, 2010/2011 and 2011/2012 school years, interactions between the year dummies and the treatment dummy, age, age squared and child fixed-effects.

Source: School performance data.

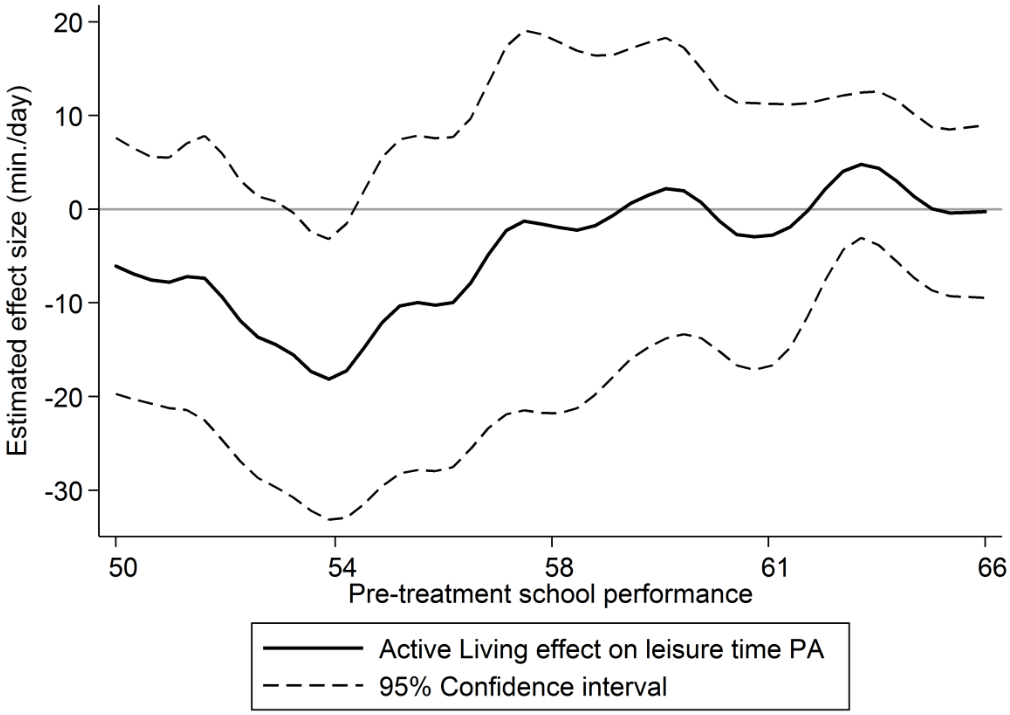
Figure A5. The effect of Active Living on time spent on physical activity during school time, by pre-treatment school performance



Notes: Local polynomial smooth graph of the results of ordinary least squares regression analyses across the pre-treatment overall school performance distribution in 2012/2013. Pre-treatment school performance is a moving window of 200 child IDs. The dependent variable in each regression is time spent on physical activity during school time. The estimated effect size based on difference-in-differences estimations is plotted on the y-axis. Independent variables are a year dummy, wearing time of the accelerometer, wearing time squared, weather variables and child fixed-effects. Adjusted 95% confidence intervals for estimated coefficients with robust standard errors are calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution, because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure A4).

Source: Merged school performance and physical activity data.

Figure A6. The effect of Active Living on time spent on physical activity during leisure time, by pre-treatment school performance



Notes: Local polynomial smooth graphs of the results of ordinary least squares regression analyses across the pre-treatment overall school performance distribution. Pre-treatment school performance is a moving window of 200 child IDs. The dependent variable in each regression is time spent on physical activity during leisure time. The estimated effect sizes based on difference-in-differences estimations are plotted on the y-axes. Independent variables are a year dummy, wearing time of the accelerometer, wearing time squared, weather variables and child fixed-effects. Adjusted 95% confidence intervals for estimated coefficients with robust standard errors are calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors. The results in the right tail of the distribution should be interpreted with caution, because of the differences in pre-treatment trends in school performance in this part of the distribution (see Figure A4).

Source: Merged school performance and physical activity data.

Table A1. Description of implemented interventions

	Description
Panel A. Active transportation to school	
Development safe route to school	A marked route to school without dangerous crossings or with guided crossings
Mobilize crossing guards	Crossing guards at the nearest (busy) road in the school environment, approximately 15 minutes before and after school hours.
Adapt unsafe intersection in school environment	Establish a new priority situation at intersections.
Availability of bicycle racks	Provide bicycle racks at schools where they were not (sufficiently) available.
School pedestrian crossing indicators	Put signs where to cross (busy) roads on foot.
Create safer parking situation around school	Redesign parking lots into parallel parking lots.
Create traffic circle in schoolyard environment	Introduce one-way streets.
Sticker competition for active transport	Provide stickers to children travelling on foot or who went to school by bike.
School stimulation documentation on safe active transport	Develop an active school transportation policy.
Introduction 'Walk/Bike-to-school-day'	Mark a day in the week or month as active transportation-day.
Speed check action performed by children	Children check speed or car users with the help of local police and provide 'feedback' to drivers.
Lessons to improve bicycle skills	Several skill lessons how to drive a bike safely.
Panel B. PA in school	
New fixed equipment in schoolyard	e.g. soccer goals, a climbing structure.
New loose equipment in schoolyard	e.g. balls, ropes.
Playground markings	e.g. paintings, hopscotch tracks.
Establish ball game area	e.g. markings as indicator of a ball games area, regulations where to play ball games.
Put a ball backstop besides railway	(Re)facilitate the use of balls at the schoolyard.
Sound equipment in schoolyard environment	Music equipment to facilitate dancing at the schoolyard.
Additional sports day in schoolyard	Once or twice per schoolyear.
Sports clinics in recess	Multiple clinics provided by local sports clubs during recess.
Use of schoolyard games	Media cards that 'instruct' children how to play games.
Prize contest for best idea PA stimulation children	Create social support to physical activity promotion at schools, e.g. walking-to-school-day.
Panel C. PA in leisure time	
Establish training circuit	A 'green' athletics track next to the schoolyard.
Active Living Games	An additional sport day for all participating school, which the children were able to 'prepare' for.
Establish out-of-school PA program	Multiple sports disciplines provided directly after school, e.g. gymnastics, athletics.
Establish school soccer team	Enhance social support in leisure time. A school team structurally plays soccer after school. Available for all interested children, irrespective their gender.
Establish PA activities by children for local residents	Establishment of 'Neighborhood in Action' group that stimulates physical activity by performing family (physical) activities, such as walks through parks .

Table A2. Differences between children in treatment and control schools in the pre-treatment year (2012/2013)

	Treatment		Control		Diff.	P-value
	mean	sd	mean	sd		
Panel A. School performance data						
Number of tests taken	4.92	(0.69)	5.09	(1.12)	-0.17	0.003
Overall school performance	57.67	(7.73)	56.86	(8.06)	0.81	0.103
Language performance	57.31	(7.77)	56.63	(8.02)	0.68	0.169
Math performance	56.49	(7.72)	55.56	(8.27)	0.93	0.065
Age when tests were taken (months)	125.51	(8.05)	126.75	(8.63)	-1.24	0.018
Gender (1=boy)	0.48	(0.50)	0.47	(0.50)	0.00	0.946
Cohort (1=5th graders, 0=4th graders)	0.55	(0.50)	0.60	(0.49)	-0.05	0.095
Panel B. PA data						
Number of days the accelerometer was worn	3.78	(1.32)	4.07	(1.27)	-0.28	0.000
Time spent on PA during school time (min./day)	85.29	(25.14)	100.28	(25.44)	-14.99	0.000
Time spent on PA during leisure time (min./day)	201.92	(52.33)	200.51	(47.24)	1.41	0.712
Age during PA measurement (months)	128.14	(8.21)	128.47	(8.27)	-0.33	0.468
Gender (1=boy)	0.49	(0.50)	0.41	(0.49)	0.08	0.002
Cohort (1=5th graders, 0=4th graders)	0.47	(0.50)	0.50	(0.50)	-0.03	0.266
Wearing time of accelerometer (min./day)	419.87	(137.63)	429.56	(128.12)	-9.69	0.176

Notes: Results from t-tests on the equality of means. Tests taken between April 3rd and August 11th 2013 are excluded from the analyses.

Sources: School performance data and PA data.

Table A3. Pre-treatment trends in overall school performance

Dep. var. : overall school performance	(1) OLS	(2) FE	(3) FE
Treatment * 2012/2013	Ref.		
Treatment * 2011/2012	-0.812	-0.734	-0.618
	(0.555)	(0.552)	(0.578)
Treatment * 2010/2011	-0.473	-0.355	-0.125
	(0.619)	(0.633)	(0.607)
Treatment * 2009/2010	-1.267	-1.010	-0.650
	(0.880)	(0.895)	(0.856)
2012/2013	Ref.		
2011/2012	-6.079	-6.131	-7.381
	(0.414)	(0.425)	(5.185)
2010/2011	-14.437	-14.545	-15.855
	(0.536)	(0.551)	(11.597)
2009/2010	-25.125	-25.372	-25.363
	(0.580)	(0.624)	(17.900)
Treatment (1=Treatment group)	0.809		
	(0.847)		
Age when tests were taken (months)			1.187
			(0.619)
Age squared			-0.005
			(0.001)
Child fixed-effects	No	Yes	Yes
Constant	56.860		
	(0.532)		
Observations	3,904	3,885	3,885
R-squared	0.580	0.910	0.922
Number of child ID		996	996
F-test Treatment*year interactions [p-value]	[0.123]	[0.162]	[0.280]

Notes: Results of three ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Tests taken between April 3rd and August 11th 2013 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. P-values for F-tests for joint significance of the Treatment*year interactions are reported in brackets.

Source: School performance data.

Table A4. Descriptive statistics for Active Living and non-Active Living schools

	Full sample		Non-Active Living schools		Active Living schools		Diff.	P-value
	mean	(sd)	mean	(sd)	mean	(sd)		
Child enrolled in an Active Living school (1=yes)	0.17	(0.38)	0.00	(0.00)	1.00	(0.00)	-1.00	0.000
Overall school performance	62.28	(9.30)	62.41	(9.33)	61.65	(9.11)	0.76	0.001
Language performance	61.62	(9.25)	61.74	(9.29)	61.04	(9.01)	0.70	0.002
Math performance	60.96	(8.88)	61.10	(8.88)	60.30	(8.86)	0.80	0.000
Age when tests were taken (months)	132.39	(10.25)	132.37	(10.25)	132.46	(10.26)	-0.09	0.718
Gender (1=boy)	0.49	(0.50)	0.49	(0.50)	0.48	(0.50)	0.02	0.189
Cohort (1=5th graders, 0=4th graders)	0.61	(0.49)	0.61	(0.49)	0.59	(0.49)	0.02	0.129
Number of child ID	6,776		5,671		1,105			
Number of schools	174		153		21			

Notes: Descriptive statistics are based on the pre-treatment year (2012/2013) and the post-treatment year (2013/2014) combined.
Source: School performance data

Table A5. Descriptive statistics for the estimation samples

	Pre-treatment year		Post-treatment year		Diff.	P-value
	mean	(sd)	mean	(sd)		
Panel A. School performance data						
Treatment (1=Treatment group)	0.52	(0.50)	0.52	(0.50)	0.00	1.000
Number of tests taken	5.00	(0.93)	4.98	(0.77)	-0.02	0.515
Overall school performance	57.28	(7.90)	65.16	(8.04)	7.88	0.000
Language performance	56.99	(7.89)	64.29	(8.09)	7.31	0.000
Math performance	56.05	(7.99)	63.81	(7.64)	7.76	0.000
Age when tests were taken (months)	126.10	(8.35)	137.91	(8.21)	11.81	0.000
Gender (1=boy)	0.48	(0.50)	0.48	(0.50)	0.00	1.000
Cohort (1=5th graders, 0=4th graders)	0.57	(0.49)	0.57	(0.49)	0.00	1.000
Number of child ID	1,014		1,014			
Panel B. PA data						
Treatment (1=Treatment group)	0.56	(0.50)	0.56	(0.50)	0.00	1.000
Number of days the accelerometer was worn	4.10	(1.15)	4.15	(1.46)	0.05	0.547
Time spent on school time PA	93.36	(24.74)	93.64	(25.64)	0.28	0.854
Time spent on leisure time PA	203.50	(47.13)	182.46	(43.99)	-21.04	0.000
Wearing time of accelerometer during school time (min./day)	309.23	(32.18)	323.73	(31.67)	14.51	0.000
Wearing time of accelerometer during leisure time (min./day)	547.89	71.55	527.24	69.71	-20.65	0.000
Age during PA measurement (months)	127.75	(8.22)	139.70	(8.23)	11.96	0.000
Gender (1=boy)	0.44	(0.50)	0.44	(0.50)	0.00	1.000
Cohort (1=5th graders, 0=4th graders)	0.45	(0.50)	0.45	(0.50)	0.00	1.000
Outside temperature during PA measurement (Celsius)	13.07	(5.58)	17.04	(3.45)	3.97	0.000
Amount of sunshine during PA measurement week (hours/day)	2.62	(1.59)	5.20	(2.26)	2.58	0.000
Rain during PA measurement week (1=yes)	0.46	(0.50)	0.47	(0.50)	0.01	0.669
Number of child ID with school time PA data	536		536			
Number of child ID with leisure time PA data	509		509			
Panel C. Merged data						
Treatment (1=Treatment group)	0.60	(0.49)	0.60	(0.49)	0.00	1.000
Overall school performance	57.67	(7.94)	65.58	(8.14)	7.91	0.000
Language performance	57.38	(7.92)	64.73	(8.23)	7.35	0.000
Math performance	56.24	(7.76)	64.09	(7.48)	7.85	0.000
Time spent on school time PA	91.84	(23.65)	92.12	(25.02)	0.28	0.867
Time spent on leisure time PA	202.65	(47.78)	182.24	(44.81)	-20.40	0.000
Wearing time of accelerometer during school time (min.day)	308.75	(31.47)	325.62	(32.40)	16.87	0.000
Wearing time of accelerometer during leisure time (min.day)	546.82	(70.43)	529.76	(69.61)	-17.06	0.001
Age when tests were taken (months)	124.89	(8.27)	136.74	(8.11)	11.85	0.000
Age during PA measurement (months)	128.36	(8.34)	140.31	(8.34)	11.95	0.000
Gender (1=boy)	0.42	(0.49)	0.42	(0.49)	0.00	1.000
Cohort (1=5th graders, 0=4th graders)	0.52	(0.50)	0.52	(0.50)	0.00	1.000
Outside temperature during PA measurement (Celsius)	12.46	(5.72)	16.80	(3.36)	4.34	0.000
Amount of sunshine during PA measurement week (hours/day)	2.59	(1.62)	4.96	(2.39)	2.37	0.000
Rain during PA measurement week (1=yes)	0.44	(0.50)	0.51	(0.50)	0.07	0.039
Number of child ID with school performance and school time PA data	422		422			
Number of child ID with school performance and leisure time PA data	398		398			

Notes: Descriptive statistics for the school performance data are based on the estimation sample of the analyses presented in column 3 of Table 3. Descriptive statistics for the PA data are based on the estimation sample of the analyses presented in columns 3 and 6 of Table 6. Descriptive statistics for the merged data are based on the estimation sample of the analyses presented in Figure 3.

Table A6. Difference-in-differences estimations of the effect of Active Living on school performance on language and math tests

Dep. var. : school performance on language or math tests	(1)	(2)	(3)	(4)	(5)	(6)
		Language			Math	
	OLS	FE	FE	OLS	FE	FE
Treatment * year	-1.231	-1.004	-1.072	-1.191	-1.216	-1.309
	(0.595)	(0.575)	(0.586)	(0.568)	(0.549)	(0.511)
Adjusted p-value	[0.055]	[0.099]	[0.087]	[0.052]	[0.041]	[0.022]
Treatment (1=Treatment group)	0.682			0.929		
	(0.821)			(0.859)		
Year (1=post-treatment year, 0=pre-treatment year)	7.979	7.832	3.892	8.409	8.444	-0.448
	(0.479)	(0.459)	(3.844)	(0.480)	(0.450)	(3.442)
Age when tests were taken (months)			0.803			1.250
			(0.418)			(0.285)
Age squared			-0.002			-0.002
			(0.001)			(0.001)
Child fixed-effects	No	Yes	Yes	No	Yes	Yes
Constant	56.630			55.561		
	(0.574)			(0.455)		
Observations	2,064	2,028	2,028	2,057	2,016	2,016
R-squared	0.175	0.751	0.754	0.200	0.783	0.789
Number of child ID		1,014	1,014		1,008	1,008
Mean school performance in estimation sample	60.722	60.640	60.640	60.015	59.938	59.938
Standard deviation	(8.814)	(8.784)	(8.784)	(8.731)	(8.724)	(8.724)
Mean school performance in full sample	61.623	61.623	61.623	60.964	60.964	60.964
Standard deviation	(9.248)	(9.248)	(9.248)	(8.882)	(8.882)	(8.882)

Notes: Results of six ordinary least squares regression analyses. The dependent variable in columns (1)-(3) is school performance on language tests. The dependent variable in columns (4)-(6) is school performance on math tests. Tests taken between April 3rd and August 11th 2013 and between March 26th and August 14th 2014 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.

Source: School performance data.

Table A7. Pre-treatment trends in school performance on language and math tests

Dep. var. : school performance on language or math tests	(1)	(2)	(3)	(4)	(5)	(6)
		Language			Math	
	OLS	FE	FE	OLS	FE	FE
Treatment * 2012/2013	Ref.			Ref.		
Treatment * 2011/2012	-0.566	-0.508	-0.415	-1.172	-1.179	-1.064
	(0.569)	(0.567)	(0.611)	(0.573)	(0.543)	(0.548)
Treatment * 2010/2011	-0.250	-0.140	0.067	-0.981	-1.006	-0.799
	(0.591)	(0.605)	(0.629)	(0.897)	(0.904)	(0.840)
Treatment * 2009/2010	-0.934	-0.676	-0.336	-1.789	-1.702	-1.386
	(0.973)	(0.983)	(0.956)	(0.812)	(0.814)	(0.784)
2012/2013	Ref.			Ref.		
2011/2012	-5.740	-5.784	-7.541	-5.786	-5.773	-6.548
	(0.439)	(0.453)	(6.148)	(0.291)	(0.259)	(2.694)
2010/2011	-13.270	-13.376	-15.873	-14.458	-14.446	-14.891
	(0.539)	(0.550)	(13.776)	(0.608)	(0.611)	(5.945)
2009/2010	-23.350	-23.609	-25.509	-24.454	-24.536	-23.467
	(0.646)	(0.680)	(21.304)	(0.619)	(0.614)	(8.954)
Treatment (1=Treatment group)	0.682			0.929		
	(0.823)			(0.860)		
Age when tests were taken (months)			1.074			1.059
			(0.736)			(0.314)
Age squared			-0.005			-0.005
			(0.001)			(0.000)
Child fixed-effects	No	Yes	Yes	No	Yes	Yes
Constant	56.630			55.561		
	(0.575)			(0.456)		
Observations	3,904	3,885	3,885	3,894	3,874	3,874
R-squared	0.536	0.881	0.893	0.565	0.898	0.907
Number of child ID		996	996		995	995
F-test Treatment*year interactions [p-value]	[0.392]	[0.430]	[0.480]	[0.042]	[0.029]	[0.097]

Notes: Results of six ordinary least squares regression analyses. The dependent variable in columns (1)-(3) is school performance on language tests. The dependent variable in columns (4)-(6) is school performance on math tests. Tests taken between April 3rd and August 11th 2013 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. P-values for F-tests for joint significance of the Treatment*year interactions are reported in brackets.

Source: School performance data.

Table A8. Difference-in-differences estimations of the effect of Active Living on overall school performance – one-by-one deletion of treatment schools

Dep. var. : overall school performance	(1) Excl. T1	(2) Excl. T2	(3) Excl. T3	(4) Excl. T4	(5) Excl. T5	(6) Excl. T6	(7) Excl. T7	(8) Excl. T8	(9) Excl. T9
Treatment * year	-1.052	-1.161	-1.075	-1.096	-1.325	-1.314	-1.111	-1.329	-1.083
	(0.543)	(0.556)	(0.562)	(0.547)	(0.561)	(0.555)	(0.553)	(0.550)	(0.550)
Adjusted p-value	[0.073]	[0.056]	[0.076]	[0.065]	[0.033]	[0.033]	[0.064]	[0.030]	[0.069]
Year (1=post-treatment year, 0=pre-treatment year)	2.632	2.006	4.334	3.174	2.517	3.016	1.514	3.418	2.373
	(3.996)	(3.972)	(4.240)	(3.853)	(3.907)	(3.848)	(3.490)	(3.821)	(3.824)
Age when tests were taken (months)	1.045	1.085	0.881	0.938	0.949	0.954	0.980	0.993	0.966
	(0.416)	(0.411)	(0.452)	(0.407)	(0.448)	(0.408)	(0.404)	(0.413)	(0.401)
Age squared	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.001	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Child fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,904	1,958	1,896	1,978	1,826	1,874	1,918	1,898	1,942
R-squared	0.813	0.813	0.815	0.814	0.807	0.810	0.823	0.811	0.814
Number of child ID	952	979	948	989	913	937	959	949	971
Mean school performance in estimation sample	61.042	61.245	61.457	61.174	61.267	61.249	61.226	61.179	61.092
Standard deviation	(8.911)	(8.921)	(8.814)	(8.898)	(8.847)	(8.888)	(8.947)	(8.863)	(8.980)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: Results of nine ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Tn = Treatment school (number) which is excluded from the analysis. Tests taken between April 3rd and August 11th 2013 and between March 26th and August 14th 2014 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.

Source: School performance data.

Table A9.Difference-in-differences estimations of the effect of Active Living on overall school performance – one-by-one deletion of control schools

Dep. var. : overall school performance	(1) Excl. C1	(2) Excl. C2	(3) Excl. C3	(4) Excl. C4	(5) Excl. C5	(6) Excl. C6	(7) Excl. C7	(8) Excl. C8	(9) Excl. C9	(10) Excl. C10	(11) Excl. C11
Treatment * year	-1.142	-1.294	-0.890	-1.145	-1.158	-1.044	-1.122	-1.373	-1.170	-1.515	-1.051
	(0.589)	(0.538)	(0.508)	(0.562)	(0.575)	(0.564)	(0.621)	(0.564)	(0.544)	(0.492)	(0.571)
Adjusted p-value	[0.073]	[0.030]	[0.102]	[0.061]	[0.064]	[0.085]	[0.092]	[0.029]	[0.048]	[0.008]	[0.087]
Year (1=post-treatment year, 0=pre-treatment year)	3.024	2.298	2.197	3.324	2.859	4.497	2.571	6.034	2.950	1.782	2.987
	(3.919)	(3.806)	(4.471)	(4.077)	(3.814)	(3.773)	(3.938)	(3.234)	(3.798)	(4.168)	(3.741)
Age when tests were taken (months)	1.089	0.997	0.865	0.983	1.013	0.923	0.927	0.686	0.972	1.217	0.955
	(0.412)	(0.409)	(0.428)	(0.410)	(0.407)	(0.409)	(0.417)	(0.344)	(0.400)	(0.408)	(0.400)
Age squared	-0.002	-0.002	-0.001	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Child fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,944	2,012	1,892	1,988	1,958	1,950	1,874	1,892	2,028	1,868	1,932
R-squared	0.809	0.819	0.814	0.810	0.809	0.809	0.804	0.816	0.812	0.820	0.807
Number of child ID	972	1,006	946	994	979	975	937	946	1,014	934	966
Mean school performance in estimation sample	61.170	61.206	61.133	61.214	61.319	61.123	61.258	61.300	61.222	61.183	61.379
Standard deviation	(8.883)	(8.914)	(8.712)	(8.901)	(8.855)	(8.895)	(8.952)	(8.918)	(8.890)	(8.982)	(8.805)
Mean school performance in full sample	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280	62.280
Standard deviation	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)	(9.296)

Notes: Results of eleven ordinary least squares regression analyses. The dependent variable in each column is overall school performance (standardized). Cn = Control school (number) which is excluded from the analysis. Tests taken between April 3rd and August 11th 2013 and between March 26th and August 14th 2014 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. Adjusted p-values for estimated coefficients with robust standard errors, calculated using a t-distribution with degrees of freedom equal to the number of groups minus the number of regressors, are reported in brackets.

Source: School performance data.

Table A10. Pre-treatment trends in school performance of children in Active Living schools compared to children in non-Active Living schools

Dep. var. : overall school performance	(1)	(2)	(3)	(4)	(5)	(6)
		Control			Treatment	
	OLS	FE	FE	OLS	FE	FE
Control * 2012/2013	Ref.					
Control * 2011/2012	0.455	0.488	0.433			
	(0.424)	(0.434)	(0.429)			
Control * 2010/2011	0.429	0.514	0.376			
	(0.556)	(0.568)	(0.517)			
Control * 2009/2010	0.540	0.579	0.326			
	(0.645)	(0.680)	(0.610)			
Treatment * 2012/2013				Ref.		
Treatment * 2011/2012				-0.357	-0.246	-0.248
				(0.384)	(0.367)	(0.394)
Treatment * 2010/2011				-0.044	0.160	0.187
				(0.356)	(0.353)	(0.363)
Treatment * 2009/2010				-0.727	-0.432	-0.398
				(0.716)	(0.697)	(0.672)
2012/2013	Ref.			Ref.		
2011/2012	-6.534	-6.619	-10.077	-6.534	-6.619	-8.041
	(0.129)	(0.128)	(1.617)	(0.129)	(0.128)	(1.798)
2010/2011	-14.866	-15.059	-21.473	-14.866	-15.059	-17.021
	(0.187)	(0.181)	(3.583)	(0.187)	(0.181)	(3.980)
2009/2010	-25.665	-25.951	-34.060	-25.665	-25.951	-27.247
	(0.310)	(0.302)	(5.474)	(0.310)	(0.302)	(6.092)
Control (1=Control group, 0=non-Active Living)	-1.056					
	(0.607)					
Treatment (1=Treatment group, 0=non-Active Living)				-0.246		
				(0.715)		
Age when tests were taken (months)			0.729			0.942
			(0.190)			(0.212)
Age squared			-0.004			-0.004
			(0.000)			(0.000)
Child fixed-effects	No	Yes	Yes	No	Yes	Yes
Constant	57.916			57.916		
	(0.313)			(0.313)		
Observations	19,894	19,765	19,759	20,100	19,976	19,970
R-squared	0.550	0.906	0.915	0.554	0.909	0.917
Number of child ID		5,075	5,074		5,123	5,122
F-test Control*year interactions [p-value]	[0.506]	[0.517]	[0.660]			
F-test Treatment*year interactions [p-value]				[0.244]	[0.220]	[0.160]

Notes: Results of six ordinary least squares regression analyses. The dependent variable in each column is overall school performance. Tests taken between April 3rd and August 11th 2013 are excluded from the analyses. Robust standard errors, clustered at the school level, are reported in parentheses. P-values for F-tests for joint significance of the Control*year and Treatment*year interactions are reported in brackets.

Source: School performance data.